A programme for bidirectional phonology and phonetics and their acquisition and evolution

by Paul Boersma, September 19, 2006
(References that supplement the information in this paper are given in **bold**.)

This paper (together with its five supplements) presents an Optimality-Theoretic (OT) grammar model that is intended to be capable of handling ‘all’ of phonology: its representations with their relations, its processes with their relations, its acquisition by the child, its evolution over the generations, and its typology across languages. The goal of the model is to achieve explanatory adequacy by doing *whole-language simulations* of the acquisition and evolution of a language.

I start by giving you the whole grammar model, then zoom in on smaller parts of it, starting with the very small.

Figure 1, then, shows the proposed minimal but comprehensive model of grammar. It is comprehensive in the sense that it is meant to be able to handle ‘all’ phonological and related phenomena. And it is ‘minimal’ in the sense that it contains what I think is the minimum number of representations that we need to do interesting phonology, namely two phonological representations that are connected to each other and to two semantic and two phonetic representations.

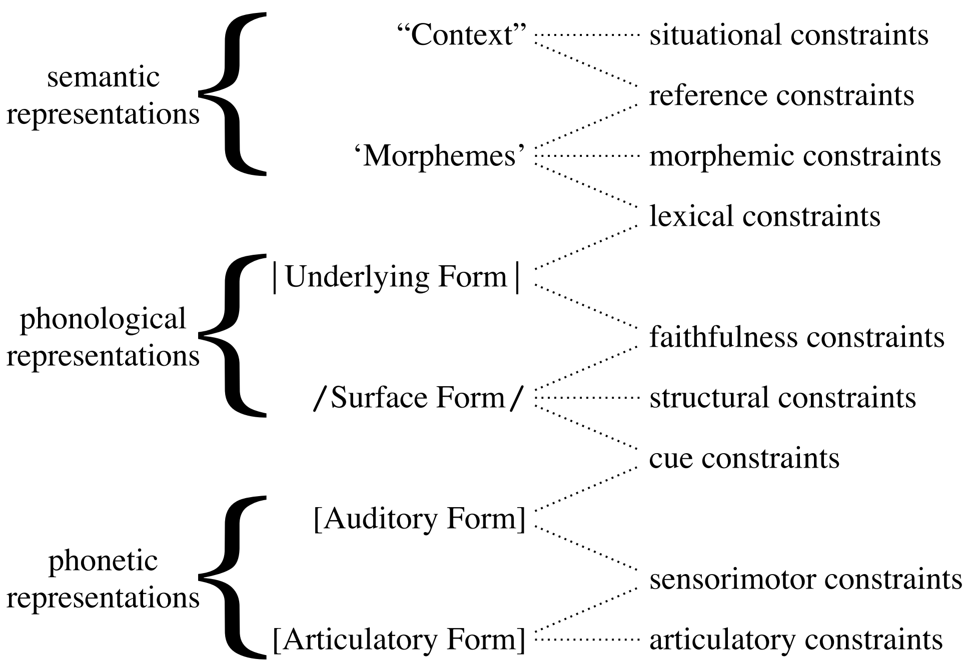

Fig. 1 The grammar model.

1. Phonological representations: Underlying and Surface Form

The minimum number of phonological representations capable of handling any interesting phonological phenomena seems to be two: at the very least we seem to require the traditional distinction between Underlying Form and Surface Form. The **Underlying Form** is usually regarded as a sequence of pieces of lexical phonological material with discernible morpheme structure, for example |an+pa|, where '+' is a morpheme boundary. The **Surface Form** is a structure of abstract phonological elements such as features, segments, syllables, and feet, for instance /.am.pa./, where '.' is a syllable boundary. For these two representations, the following subsections describe their relations, their role in merely-phonological processes, and their role in merely-phonological acquisition and evolution.

1.1. The relation between Underlying Form and Surface Form

From Prince & Smolensky (1993) on, the relation between these two representations has been modelled in terms of the **faithfulness constraints** that are visible in Figure 1. For instance, the combination of the underlying form |an+pa| and the surface form /.am.pa./ constitutes a *faithfulness violation* because the underlying |n| corresponds to a surface /m/ that is not identical to it (McCarthy & Prince 1995).

1.2. The process of merely-phonological production

Prince & Smolensky, and most OT-ists following them, have regarded phonology as being primarily concerned with the unidirectional process of *phonological production*, i.e. the mapping *from* Underlying Form *to* Surface Form, as in Figure 2.

|Underlying Form|

/Surface Form/

Fig. 2 The merely-phonological production process.

In this mapping, the faithfulness constraints can interact with the **structural constraints** that are also visible in Figure 1. For instance, the mapping from underlying |an+pa| to surface /.am.pa./ could be due to a constraint against codas that do not share their place with a following onset. In order to force the surfacing of /.am.pa./, this constraint has to outrank the faithfulness constraint mentioned in the previous paragraph. Tableau (1) shows how this works in OT.[1] The constraint names contain subscripts for the representations that they evaluate (S for Surface, US for Underlying & Surface).

[1] For readers not familiar with OT tableaus: the top left cell is the 'input', the cells below it are candidates; the remaining columns show the constraints ordered by their ranking; an asterisk denotes a constraint violation, an exclamation mark a fatal violation (one that outlaws a candidate). The finger points at the *winning candidate*. This candidate is also called the *optimal output*, because it minimally violates the ranked constraints.

(1) *Phonological production without semantics or phonetics*

|an+pa|	*CodaWithSeparatePlaces$_{S}$	IdentPlace$_{US}$
|an+pa| /.an.pa./	*!	
☞ |an+pa| /.am.pa./		*
|an+pa| /.aŋ.pa./	*!	*

The notation in tableau (1) is slightly different from the usual notation in that the candidate cells contain paired representations, i.e., the ‘input’ (the underlying form |an+pa|) has been included in each cell. This manner of writing candidates will become especially relevant when we consider cases with more than two representations, and cases with bidirectional learning tableaus.

The interpretation of tableaus like (1) is that the candidates listed are all those that share the representations in the top left cell. Thus, Gen (the OT candidate generator) generates three UF–SF pairs that contain the representation |an+pa|, and these three are the candidates in (1).[2] For the evaluation of the constraints, we have to look only at the candidates themselves. For instance, the violation of IdentPlace$_{US}$ by the second candidate can be detected by sole inspection of the pair |an+pa| /.am.pa./, without comparing the candidate to the ‘input’ |an+pa| separately.

1.3. The process of merely-phonological comprehension

The OT model with two representations (Underlying and Surface Form) has been used to a limited extent bidirectionally, namely to model *phonological comprehension* besides phonological production, as in Figure 3.

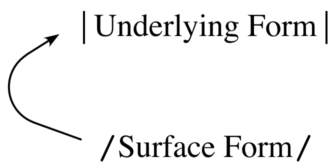

Fig. 3 The merely-phonological comprehension process.

Smolensky (1996) mentions the case of a learner with a high-ranked structural constraint, who cannot produce all forms that she can comprehend. Smolensky’s argument can be expressed with the same example as above: the ranking tableau (1) can be regarded as the grammar of a child who cannot yet produce /.an.pa./. If we suppose that adults of the same language have the reverse ranking, these adults must be able to produce /.an.pa./, and this form will occur in the learner’s environment. Tableau (2) shows that the learner successfully comprehends this form.

[2] In an example with more faithfulness constraints that can be violated, there will be more than three candidates.

(2) *Phonological comprehension without semantics or phonetics*

/.an.pa./	*CODAWITHSEPARATEPLACES	IDENTPLACEUS
☞ |an+pa| /.an.pa./	*	
|am+pa| /.an.pa./	*	*!
|aŋ+pa| /.an.pa./	*	*!

The candidates in (2) are now all the thinkable paired representations that share the Surface Form /.an.pa./. Crucially in (2), all of these doublets violate the structural constraint with the long name, so that the decision falls to the faithfulness constraint. This is how Smolensky solved what he called the ‘production-comprehension dilemma’, the case where a learner’s comprehension skills precede her production skills.

1.4. Merely-phonological acquisition

A **bidirectional learning tableau** is a tableau with two input representations rather than just one. It helps in improving the relation between the two representations. For the case of mere phonology, it helps to improve both of the mappings in Figure 4.

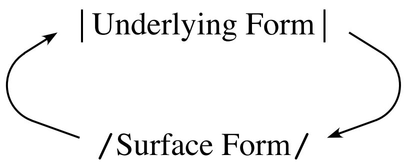

Fig. 4 Bidirectional acquisition of mere phonology.

In tableau (3), the assimilating learner of tableau (2) has obtained knowledge that the underlying form |an+pa| and the surface form /.an.pa./ can go together (i.e. the parents do not assimilate). This pair is therefore included in the top left cell of the tableau. The candidates in the tableau are all the underlying-surface pairs that share one or two of the representations, i.e. all the possible pairs that contain either |an+pa| or /.an.pa./, or both.

(3) *Phonological acquisition without semantics or phonetics*

|an+pa| /.an.pa./	*CODAWITHSEPARATEPLACES	IDENTPLACEUS
√ ☜ |an+pa| /.an.pa./	*→	
☞ |an+pa| /.am.pa./		←*
|an+pa| /.aŋ.pa./	*	*
|am+pa| /.an.pa./	*	*
|aŋ+pa| /.an.pa./	*	*

The interpretation of the pair |an+pa| /.an.pa./ in the top left cell is as follows. The learner has just heard the surface form /.an.pa/, and successfully comprehended it. Something (for instance the discourse context) subsequently tells her that the correct underlying form that corresponds to this instance of the surface form /.an.pa/ was |an+pa|. On the basis of this information the learner regards |an+pa| /.an.pa./ as a correct combination of underlying and surface form.

The interpretation of the three marks in the candidate cells in (3) is as follows. The backwards-pointing finger ("☜") marks the candidate that would win in the comprehension direction, if only the 'correct' surface form /.an.pa./ were given. This form is shared by candidates 1, 4, and 5, and of these three candidates candidate 1 is the most harmonic given the constraint ranking. The forward-pointing finger ("☞") marks the candidate that would win in the production direction if only the 'correct' underlying form |an+pa| were given. This form is shared by candidates 1, 2, and 3, and candidate 2 is the most harmonic of these. The check mark ("√"), finally, marks the most harmonic of all the candidates that share both the underlying form |an+pa| and the surface form /.an.pa./. There is obviously only one candidate that shares both forms, namely candidate 1, so candidate 1 is immediately the most harmonic of all such candidates.

The interpretation of the arrows in (3) is as follows. The candidate with the check mark is regarded by the learner as the *correct candidate*. The *forward winner* ("☞") differs from this correct candidate, so the learner has evidence that the forward winner is an incorrect pair. As a result, the learner will take action by taking a *learning step*: she raises the rankings of all the constraints that are violated in the incorrect forward winner (in this case, only IDENTPLACE$_{US}$), and lowers the rankings of all the constraints that are violated in the correct candidate (in this case, only *CODAWITHSEPARATEPLACES$_{S}$). These raisings and lowerings are depicted with arrows in the tableau. With 'raising' and 'lowering' I mean that these constraints move a small distance along the ranking scale of Stochastic OT (Boersma 1997, Boersma & Hayes 2001). These movements make it thereby more likely that a future |an+pa| will be produced as /.an.pa./. This half of the bidirectional learning procedure is identical to the application of the Gradual Learning Algorithm (Boersma 1997) to merely-phonological cases (Boersma & Hayes 2001).

The other half of the bidirectional learning procedure happens if the *backward winner* ("☜") differs from the 'correct' candidate. This does not happen in tableau (3), but tableau (4) shows a case. This tableau shows a learner who performs no place assimilation, although place assimilation does occur in her environment, as indicated by the learning pair |an+pa| /.am.pa./.

(4) *Phonological acquisition without semantics or phonetics*

|an+pa| /.am.pa./	IDENTPLACE$_{US}$	*CODAWITHSEPARATEPLACES$_{S}$	X
☞ |an+pa| /.an.pa./		←*	
√ |an+pa| /.am.pa./	*→→		
|an+pa| /.aŋ.pa./	*	*	
☜ |am+pa| /.am.pa./			←*
|aŋ+pa| /.am.pa./	*		

In tableau (4), the mismatch between the ‘correct’ candidate and the backward winner leads to a learning step analogous to the one described above: the constraints violated in the ‘correct’ candidate are lowered, as indicated in the tableau by the second right-pointing arrow, and the constraints violated by the incorrect backward winner are raised; in order to be able to make this visible, I added a mysterious constraint X to the tableau, which is violated by the fourth candidate (see §1.6 for discussion). The changes in the constraint rankings raise the likelihood that a future occurrence of /.am.pa./ will be comprehended as |an+pa|.

1.5. Merely-phonological evolution

In the examples of §1.4, and more generally in merely-phonological learning, the learner will end up in exactly the same language as her parents. That is, given a certain underlying form she and her parents will produce the same surface form, and given a certain surface form, she and her parents will comprehend the same underlying form. The language will not change over the generations.

1.6. What is wrong with merely-phonological grammars?

Sections §1.2 through §1.5 assumed that phonology lives on an island. In reality, it is connected upwards to the semantics (and the syntax, and the pragmatics) and downwards to the phonetics, and these connections are felt throughout the phonology. This sections points out problems with the production model of §1.2, the comprehension model of §1.3, the acquisition model of §1.4, and the evolution model of §1.5.

The merely-phonological production model of §1.2 has problems accounting for many observed phonological typologies, such as universal hierarchies of frequency of occurrence (an aspect of ‘markedness’), universal hierarchies of the degree of phonological activity (another aspect of ‘markedness’), and universal hierarchies of faithfulness rankings (yet another aspect of ‘markedness’). The OT literature on merely-phonological grammars has proposed innate rankings of structural constraints (Prince & Smolensky 1993), innate rankings of faithfulness constraints (Beckman 1998), and extralinguistic knowledge of auditory contrast (Steriade 1995, 2001). These proposals come with their own problems: innate rankings turn out to have exceptions wherever such exceptions would be functionally advantageous (see Steriade 1995 against positional faithfulness, and Boersma 1998 against the sonority hierarchy), and universal rankings turn out to correspond to linguistic perceptual contrast rather than to extralinguistic auditory contrast (e.g. Boersma & Escudero 2004).

The solution is to allow the phonology to interact with the phonetics. Many observed phenomena like *auditory enhancement* (Flemming 1995), *licensing by cue* (Steriade 1995), and things often attributed to innate *markedness* will fall out automatically as side effects of learning. Merely-phonological production tableaus will often no longer be valid: the choice for |an+pa| /.am.pa./ in (1), for instance, does not have to be determined by a structural constraint at all: constraints further down Figure 1, most notably the articulatory constraints, could take care of that, as shown in §6.1.

The merely-phonological comprehension model of §1.3 also has problems. Suppose that the lexicon of the listener contains the items |an|, |am|, and |aŋ|, and that each of them can be concatenated with |pa|. In an assimilating language like the one in (3), all three cases will end up as the surface structure /.am.pa./. However, when confronted with the surface form /.am.pa./, the listener has no option but to comprehend the underlying form |am+pa|, as tableau (5) shows.

(5) *A failure of phonological comprehension without semantics*

/.am.pa./	*CODAWITHSEPARATEPLACES	IDENTPLACEUS
|an+pa| /.am.pa./		*!
☞ |am+pa| /.am.pa./		
|aŋ+pa| /.am.pa./		*!

This problem with Smolensky's (1996) proposal was first noted by Hale & Reiss (1998). The solution they proposed was that comprehension is not handled by tableaus like (5) but instead follows a procedure that yields a list of underlying forms that produce the same surface form, In the present case, all three candidate underlying forms (|an+pa|, |am+pa|, and |aŋ+pa|) yield the requested surface form /.am.pa./, so that all three remain as comprehension candidates, to be disambiguated higher up by syntactic, semantic, and pragmatic processing. This proposal comes with its own problems, like working with lists of candidates but not doing OT.

The solution is to allow the phonology to interact with the semantics. The choice for |an+pa| /.am.pa./ in (4) is then not entirely determined by faithfulness constraints: constraints further up Figure 1, such as the morphemic and situational constraints, could play a role. For instance, a situational constraint could say that |am+pa| is not an appropriate sequence of morphemes in the present discourse. This is also the constraint X in tableau (5). More details are in §9.1.

Next, the merely-phonological acquisition model in §1.4 has its problems. Tableau (3) only works if the learner has knowledge both of the underlying form and of the surface form. But in reality, that information is not directly available to the learner. Both forms must be based on something the learner has heard, perhaps an auditory-phonetic form such as [ampa]. From this, the learner first has to construct the abstract discrete phonological surface form /.am.pa./. This perceptual construction process has to rely on language-specific knowledge of the relation between phonetic detail and discrete phonological elements, and the process itself is language-specific and interacts with the phonology (see §3.2). The second thing the learner has to construct is the underlying form |an+pa|. Something must have told her that this form is correct, rather than a competing underlying form |am+pa| that could also be in the lexicon. The recognition process has to rely on language-specific knowledge of the relation between phonological structures and meaning, and the process itself is language-specific and interacts with the phonology (as seen above and in §9.1).

The solution is to model both the perceptual construction process and the recognition process in OT, because both are language-specific and interact with the phonology.

Finally, the merely-phonological evolution model in §1.5 has its problems. Phonological change is not predicted, although it happens in reality. Also, it cannot model the existence of **transmission noise**, which is the phenomenon that what the learner hears has been distorted by background noise. Of course, it is possible to think that this transmission noise is precisely what causes phonological change (e.g. Ohala 1981), but that can be shown to be incorrect (§4.3).

The solution is to model the phonetics in OT and have it interact with the phonology. It will turn out that automatic biases arising during acquisition will counteract the transmission noise, so that equilibria are allowed to emerge within a few generations (§4.3).

The following four sections describe the phonetic and semantic representations and their relations, and how they can interact with the phonology.

2. Phonetic representations: Auditory and Articulatory Form

The minimum number of phonetic representations seems to be two.

The **Auditory Form** is a sequence of events on auditory continua such as pitch, noise, spectral peaks and valleys, and silences, their durations, and their relations such as simultaneity and order. For instance, the *microscopic auditory transcription* [[aãm_^{p}a]] (Boersma 1998) is a shorthand notation for vocalic material with a high first formant, followed by the same but with a nasal spectral peak and valley added, followed by the spectral resonance that reflects the nasal cavity and a long oral sidebranch, followed by a silence, followed by a burst with low spectral features, followed by vocalic material with a high first formant (it is the cross-linguistically most common sound associated with the phonological structure /.am.pa./).

The **Articulatory Form** is a sequence of gestures by the multiple articulatory muscles that move, hold, tense, or relax the glottis, the larynx, the epiglottis, the pharynx walls, the tongue tip, the tongue body, the velum, the lips, the cheeks, the jaw, and the lungs. For instance, the phonetic transcription [aãmpa] is a very rough approximating shorthand for an articulation with constantly applied lung pressure and glottal adduction, starting with a lowered jaw, a low tongue, open lips, and a closed velum, followed by a lowering of the velum, followed by a closure of the lips (with jaw raising), followed by a raising of the velum, followed by an opening of the lips (with jaw lowering). This description of the Articulatory Form in terms of movements is still rather sketchy, because a description in terms of muscle activities (Boersma 1998) is much more precise.

For these two representations, the following subsections describe their relations, their role in merely-phonetic acquisition, and evolution.

2.1. The relation between Auditory Form and Articulatory form

As a speaker/listener, you have knowledge of what your articulations will sound like, and conversely of how to implement articulatorily a sound that you want to produce. This is expressed by **sensorimotor constraints**, which say such things as "an auditory high F1 does not correspond to an articulatory raised jaw."

The ranking of these constraints is less language-specific than that of other constraints, because the shapes of our vocal tracts do not depend on the language we are learning. So the rankings of these constraints, *if* they have been learned, are universal (or perhaps they depend on the speaker). Areas in auditory or articulatory space that your language does not use at all will probably lead to poor sensorimotor knowledge (variable constraint ranking) in those areas, though.

2.2. The processes of merely-phonetic audition and articulation

We are unlikely to have much time to discuss these two processes. Instead, I will usually assume perfect sensorimotor knowledge, i.e., the speakers-listeners that are modelled have a firmly fixed relation between Auditory Form and Articulatory Form, caused by a lot of learning that caused the sensorimotor constraints against incorrect auditory-articulatory relations to become high-ranked, and constraints against correct auditory-articulatory relations to become low-ranked.

2.3. Merely-phonetic acquisition

This happens every time you speak, or every time an infant tries out her speech apparatus in vocal play. Tableaus are analogous to those for mere phonology in §1.4.

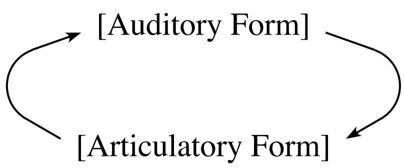

Fig. 5 Bidirectional acquisition of mere sensorimotor knowledge.

2.4. Merely-phonetic evolution

This does not exist. There is no way I can put my very personal sensorimotor knowledge into my child's head.

3. The phonology-phonetics interface

There are several theories about how phonetic representations can be connected to phonological representations. From the phonological side, the connecting representation will always be the Surface Form, never the Underlying Form, and this is illustrated in Figure 1. From the phonetic side, the situation is less clear: Figure 1 proposes that the connecting representation is the Auditory Form, but the theory of *Direct Realism* proposes that the connection is made via the Articulatory Form instead. I will assume that Figure 1 is correct, because it can be shown to work quite well. Direct Realists are invited to interchange the two phonetic representations in Figure 1 and show that that alternative grammar model works equally well or better.

3.1. The relation between Surface Form and Auditory Form

Following Figure 1, the phonetics-phonology interface is a relation expressed in terms of *cues*: auditory events in the Auditory Form can be *cues to* phonological elements in the Surface Form.

Cross-linguistically speaking, auditory cues are arbitrarily related to phonological elements. In English, the major cue to the phonological feature /voiced/ at the end of a word is the duration of the preceding vowel: the vowel of /liːdə/ 'leader' is produced longer than that of /liːtə/ 'litre', and the difference is even larger in monosyllables like /ɹoud/ 'road' and /ɹout/ 'wrote'. No such gigantic differences are found in most other languages, like e.g. between German productions of /liːdɐ/ 'songs' and /liːtɐ/ 'litre'. In production, therefore, auditory cues for voicing are used differently in German than in English.

The cross-linguistic differences in cue use are bidirectional. The differences in cue use in production are reflected in how the cues are weighted in perception. Thus, English listeners but not German listeners rely strongly on the duration of the preceding vowel when having to decide whether they heard a voiced or voiceless consonant. The same correlation between production and perception holds between closely related varieties of the same language. For instance, the differences between an /i/ and /ɪ/ produced by Scottish English lies mainly in spectral differences, namely the F1 (first formant), whereas the same phonological difference in Southern British English is implemented to a large degree in duration as well.

Clearly, the interface between phonology and phonetics is bidirectional, so that the **cue constraints** have to be formulated bidirectionally. For the case of English voiced consonants, one of the relevant cue constraints can be written as follows (the index *i* denotes correspondence; the symbol "·" means that the two adjacent elements are in the same morpheme):

(6) *A vocalic cue constraint for voicing*

*/V_i·{C,+voi}/ [–lengthened v_i]
"non-lengthened auditory vocalic material does not correspond to a phonological vowel before a tautomorphemic voiced consonant"

The formulation in (6) is not entirely correct yet, because the notation of the auditory value [–lengthened] suggests discreteness (it is either plus or minus). In general, there will be a continuous range of possible values along every auditory continuum. This is why Escudero & Boersma (2003) proposed continuous ranges of cue constraints such as those in (7).

(7) *Arbitrary cue constrains for vowel classification*

*/ɪ/ [F1 = *x* Hz]
"a first formant of x Hz does not correspond to the phonological vowel /ɪ/"

Constraints as in (7) are then thought to exist for every possible value of *x* between, say, 200 and 1200 Hz. Thus, a conspicuous property of the constraints in (7) is that they are *arbitrary*, i.e., they exist even for first formant values that are typical of the vowel /ɪ/. With arbitrary cue constraints it is the task of the constraint ranking, not the task of the constraint set, to make sure that /ɪ/ connects to plausible auditory events.

The formulations in (6) and (7) are bidirectional, e.g., (6) can be read equally well as "non-lengthened auditory vocalic material should not be perceived as a phonological vowel before a tautomorphemic voiced consonant" and as "a phonological vowel before a tautomorphemic voiced consonant should not be produced as non-lengthened auditory vocalic material." These two directions of processing are discussed in the following two subsections.

3.2. The process of prelexical perception

If the mapping from Auditory Form to Surface Form is regarded in isolation, it does not involve any processing at higher levels, most notably the lexicon. For this reason, psycholinguists call this process *prelexical perception*. Phoneticians, who tend to be less involved with matters lexical, usually stay with the term *perception*.

/Surface Form/

[Auditory Form]

Fig. 6 The prelexical perception process.

Since the relation between Auditory and Surface Form is expressed in terms of cues, an OT modelling of the mapping from Auditory to Surface Form is expected to involve cue constraints. An account of the perception of English /i/ and /ɪ/ in terms of cue constraints alone was given by Escudero & Boersma (2003, 2004). As an example, they considered the auditory event [vocalic material, F1 = 349 Hz, duration = 74 ms], to be abbreviated as [349 Hz, 74 ms]. In a Scottish English environment this event will be perceived as the vowel /i/, because 349 Hz would be a too low first formant for /ɪ/ in that variety of English:

(7) *Vowel perception in Scottish English*

[349 Hz, 74 ms]	*/ɪ/ [349 Hz]	*/i/ [74 ms]	*/ɪ/ [74 ms]	*/i/ [349 Hz]
/ɪ/ [349 Hz, 74 ms]	*!		*	
☞ /i/ [349 Hz, 74 ms]		*		*

In a Southern British English environment, the same auditory event will be perceived as /ɪ/, perhaps because it is too short to be a plausible Southern /i/:

(8) *Vowel perception in Southern British English*

[349 Hz, 74 ms]	*/i/ [74 ms]	*/i/ [349 Hz]	*/ɪ/ [74 ms]	*/ɪ/ [349 Hz]
☞ /ɪ/ [349 Hz, 74 ms]			*	*
/i/ [349 Hz, 74 ms]	*!	*		

But cue constraints are not the only constraints that pose restrictions on the outcome of prelexical perception. We can see in Figure 1 that structural constraints directly evaluate Surface Forms, so they ought to interact with cue constraints in perception. The first account of what can (with hindsight) be called perception with structural constraints in OT is that by Tesar (1997, 1998, 1999) and Tesar & Smolensky (1998, 2000). In their examples of *robust interpretive parsing*, an *overt form*, which is a string of syllables marked for stress but not for foot structure, is interpreted as a *full structural description*. The overt form [σ ˈσ σ], for instance, is a sequence of an unstressed, as stressed, and an unstressed syllable. In the left-aligning iambic language of tableau (9) this overt form is interpreted as the left-aligned iamb /(σ ˈσ) σ/.

(9) *Perception of metrical structure in a left-aligning iambic language*

[σ ˈσ σ]	FOOTBIN$_S$	FOOTLEFT$_S$	IAMBIC$_S$	TROCHAIC$_S$	FOOTRIGHT$_S$
☞ /(σ ˈσ) σ/ [σ ˈσ σ]				*	*
/σ (ˈσ σ)/ [σ ˈσ σ]		*!	*		
/σ (ˈσ) σ/ [σ ˈσ σ]	*!	*			*

In the right-aligning trochaic language of tableau (10) the same overt form is interpreted as the right-aligned trochee /σ (ˈσ σ)/.

(10) *Perception of metrical structure in a right-aligning trochaic language*

[σ ˈσ σ]	FOOTBIN$_S$	FOOTRIGHT$_S$	TROCHAIC$_S$	IAMBIC$_S$	FOOTLEFT$_S$
/(σ ˈσ) σ/ [σ ˈσ σ]		*!	*		
☞ /σ (ˈσ σ)/ [σ ˈσ σ]				*	*
/σ (ˈσ) σ/ [σ ˈσ σ]	*!	*			*

Because of the striking parallels, ‘robust interpretive parsing’ can be equated with prelexical perception, the ‘overt form’ can be regarded as a somewhat abstract variety of Auditory Form, and the ‘full structural description’ can be equated with the Surface Form (Boersma 2003); this is the interpretation assumed by Apoussidou & Boersma (2003, 2004) and the reason for the notations in (9) and (10).

The big point that Tesar and Smolensky made was that structural constraints are needed both in the production direction (§1.2) and in the comprehension direction (the metrical examples above). Hence, if these structural constraints are ranked in an OT manner and if they influence production in an OT way, then they also influence comprehension in an OT way. Hence, interpretive parsing (which is prelexical perception) should be handled in OT if phonological production is.

The example by Tesar and Smolensky does not seem to involve cue constraints, but that is only because they did not consider candidates with different numbers of syllables or different stress patterns. One can imagine cue constraints for the relation between auditory intensity, pitch and duration on the one hand and phonological stress (headship of a foot) on the other, and one can imagine that these cue constraints interact in a parallel manner with the structural constraints in (9) and (10). For instance, in an iambic language a higher pitch on the first syllable may not turn that syllable into a foot head, whereas in a trochaic language it may.

The most general cases of perception involve an interaction of structural and cue constraints. For these, see **Boersma (2006b)** on the McGurk effect, and **Boersma (2006c)** on Polivanov's idea of phonological perception.

3.3. Unidirectional acquisition of prelexical perception

I now explain the perceptual learning algorithm proposed by Boersma (1997), with an example from Escudero & Boersma (2004).

Suppose that a Scottish English child, at some point during her acquisition period, has a grammar that would be appropriate for listening to Southern British English, i.e. with the ranking of tableau (8). Now suppose that a Scottish English adult pronounces the auditory form [349 Hz, 74 ms]. As tableau (11) shows with the backward pointing finger, the child will perceive this as /ɪ/.

(11) *Vowel perception in Southern British English*

/i/ [349 Hz, 74 ms]	*/i/ [74 ms]	*/i/ [349 Hz]	*/ɪ/ [74 ms]	*/ɪ/ [349 Hz]
☜ /ɪ/ [349 Hz, 74 ms]			←*	←*
√ /i/ [349 Hz, 74 ms]	*!→	*→		

However, it is likely that the Scottish adult speaker intended the vowel /i/ instead. It is quite possible that the child will detect this. Perhaps the speaker said *please*, so that the child's lexicon, which contains an entry |pliz| but not |plɪz|, can already tell her that she should have perceived an /i/. Or the speaker intended *sheep* and the child's lexicon was satisfied with recognizing |ʃɪp| 'ship', but subsequent semantic and conceptual processing in the given situational context made her decide that the speaker had actually intended |ʃip| 'sheep'. Either way, the child will know that she has made a **perception error**, and mark the second candidate in (11) as 'correct'. As a result, the child will rerank the relevant constraints according to the arrows in (11), analogously to tableaus (3) and (4).

This procedure is called *lexicon-driven learning of perception*: the ultimately recognized Underlying Form *supervises* the learner's perception, i.e. determines what she should have perceived. This is probably an oversimplification arising from not considering the upward relations of the Surface Form in Figure 1.

A conspicuous property of tableau (11) is that it does not consider forward learning, i.e. it does not include candidates with the surface form /i/ but different auditory forms. So this is a case of unidirectional (only backward) learning. Section §3.4 has more to say about this.

3.4. The process of prototype selection

The process of prelexical perception can be reversed, as in Figure 7. This process answers the question: given the phonological Surface Form /i/, what is the best Auditory Form associated with it?

/Surface Form/

[Auditory Form]

Fig. 7 The prototype selection process.

Boersma (2005b) argued that this process cannot really be called ‘production’, because articulatory considerations are not involved. Not bound by articulatory effort, the resulting winning auditory form may well be much more ‘peripheral’ (lower F1, higher duration) than the average auditory realization. In fact, the learning algorithm described in §3.2 leads to such a situation. The idea is that the average token of /i/ may have an F1 of 330 Hz, but that 290 Hz is an even better token because it has less chance of being perceived as anything but /i/ (e.g. the typical /ɪ/ token has an F1 around 500 Hz).

(12) *Prototype selection in Scottish English*

/i/	*/i/ [74 ms]	*/i/ [349 Hz]	*/i/ [200 ms]	*/i/ [280 Hz]
/i/ [349 Hz, 74 ms]	*!	*		
/i/ [280 Hz, 74 ms]	*!			*
/i/ [349 Hz, 200 ms]		*!	*	
☞ /i/ [280 Hz, 200 ms]			*	*

For a more detailed gigantic tableau see Boersma (2005b).

3.5. Acquisition of prototype selection?

There is no separate learning algorithm for prototype selection. This is because prototype selection is just a **paralinguistic task**, i.e. you can find the effect in the laboratory (Johnson, Flemming & Wright 1993), but it is not a **linguistic task** like production and comprehension, which human evolution has optimized.

3.6. The evolution of the phonology-phonetics interface

The learning algorithm described in §3.3 is not necessarily stable over the generations. This is because the child learns to mimic the auditory frequency distributions that she hears in her environment. When she becomes a parent, she will produce these same auditory distributions. But her child will not exactly hear these distributions: there will be an additional **transmission noise** caused by wind and speaker variation. For instance, the fact that Dutch listeners use duration instead of F2 as the main cue to distinguish /a/ from /ɑ/ (Gerrits 2001: 89) is because the F2, being highly regionally dependent, is less reliable in the environment. So the auditory environment is different from the cue use of any single speaker. It would seem, then, that if the child mimics this variation, she will end up having a much broader distribution of auditory values than her parents.

Fortunately, the prototype effect described in §3.4 counteracts this drift, as we will see below.

3.7. Is this how the phonology-phonetics interface works?

At least McQueen & Cutler (1997) argue that it indeed works this way, namely that prelexical perception is modular and receives no feedback from higher processing, such as from the lexicon or from the conceptual systems. Whether the auditory-surface form connection is equally detached in production remains to be seen.

4. The three ‘low’ representations: Articulatory Form – Auditory Form – Surface Form

In §3.4, we saw that Auditory Form and Surface Form are not sufficient for modelling production. We therefore now include Articulatory Form as well.

4.1. The process of phonetic production

Once we have three representations, there are two ways to get from one end to the other: serial and parallel. Figure 8 shows the serial edition of getting from Surface Form to Articulatory Form in phonetic production:

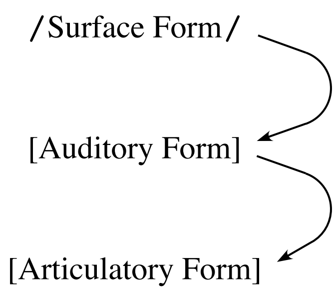

Fig. 8 The phonetic production process, serial edition.

What this means is that the speaker, given the phoneme /i/, first computes an auditory prototype (by means of **cue constraints**), say [F1=280 Hz], then turns this prototype into an articulation, which because of articulatory constraints only produces an F1 of 330 Hz. It is possible that this works. The **sensorimotor constraints** would prefer to generate an articulation that produces an F1 of 280 Hz, but the **articulatory constraints** would prevent this. This possibility has not been investigated in the literature. You are invited to compare it to the parallel model shown below.

The parallel edition looks as follows:

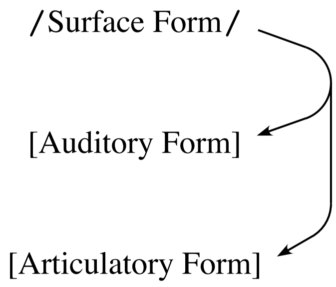

Fig. 9 The phonetic production process, parallel edition.

What this means is that the Auditory and Articulatory Forms are computed at the same time, and that the cue constraints can interact with the sensorimotor and articulatory constraints. For an example, you are referred to **Boersma & Hamann (2006)**.

4.2. Acquisition of phonetic knowledge

Boersma & Hamann (2006) use the unidirectional perceptual learning algorithm of §3.3 to show how a child learns to produce sibilants like /s/ and /ʃ/, under the assumptions that (1) phonetic production is parallel, (2) sensorimotor knowledge is perfect, and (3) articulatory constraints have a fixed ranking. Figure 10 shows the processes involved (prelexical perception and parallel phonetic production).

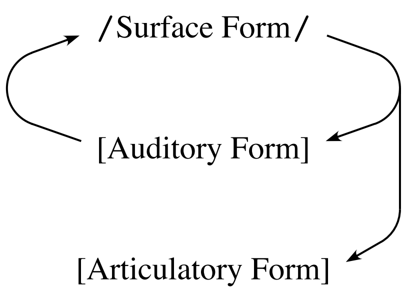

Fig. 10 The acquisition of parallel phonetic production.

Boersma & Hamann show that acquisition comes with an automatic bias towards balancing distinctivity and articulatory effort, without the assumption that the learner has any knowledge of auditory distances (hence no need for Flemming's 1995 MINDIST constraints).

4.3. Evolution of phonetic implementation

Boersma & Hamann (2006) also showed that the acquisition model in §4.2 leads to a stable evolution of the language over the generations. Even if a language starts with a skewed and rare set of sibilants, say [ç] and [s̺], it will achieve a stable equilibrium of [s] and [ʃ] within a few generations. The equilibrium is achieved when the transmission noise is exactly counterbalanced by the articulatory effort associated with extreme auditory values. An optimal balance between articulatory ease and perceptual distinctivity is thus obtained without the assumption that the learner has any knowledge of auditory distances.

5. The three 'middle' representations: Auditory Form – Surface Form – Underlying Form

We go one level up. Since this triplet of representations does not include the Articulatory Form, the only process we can handle is comprehension. This can be done serially or in parallel.

5.1. The serial edition of the process of phonetic-phonological comprehension

In a serial view of comprehension, prelexical perception is followed by word recognition, as in Figure 11.

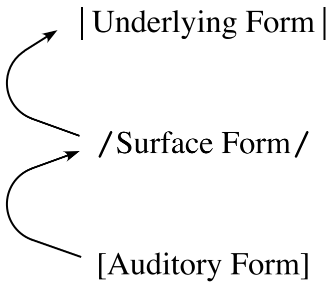

Fig. 11 The phonetic-phonological comprehension process, serial edition.

I will discuss an example. Suppose the auditory form is a sound that sounds like a typical Spanish *barte* (which is a nonsense word) or *parte* (which means 'part'), or something in between. The following tableaus ignore every auditory aspect of this sound except the voice onset time of the initial plosive.

Step one is prelexical perception, i.e. the mapping from the given Auditory Form to a phonological surface structure (Surface Form). The cue constraints are ranked by distance to the boundary, which is at –10 milliseconds. The worst token of /p/ is one with a very negative VOT such as –100 ms, so the cue constraint that says that [–100] should not be perceived as /p/ is high-ranked. Likewise, constraints that connect large positive VOT values to /b/ are also high-ranked. An appropriate ranking for perceiving Spanish voicing must be similar to that in tableaus (13) to (15).

(13) *Spanish classification of voicing*

[–100 ms]	*/p/ [–100]	*/b/ [+30]	*/p/ [–20]	*/b/ [–20]	*/p/ [+30]	*/b/ [–100]
☞ /.baɾ.te./						*
/.paɾ.te./	*!					

(14) *Spanish classification of voicing*

[–20 ms]	*/p/ [–100]	*/b/ [+30]	*/p/ [–20]	*/b/ [–20]	*/p/ [+30]	*/b/ [–100]
☞ /.baɾ.te./				*		
/.paɾ.te./			*!			

(15) *Spanish classification of voicing*

[+30 ms]	*/p/ [–100]	*/b/ [+30]	*/p/ [–20]	*/b/ [–20]	*/p/ [+30]	*/b/ [–100]
/.baɾ.te./		*!				
☞ /.paɾ.te./					*	

Step 2 is word recognition. The underlying form |paɾte| ‘part’ exists, the underlying form |baɾte| does not. The perceived form /.paɾ.te./ will easily be recognized with the help of faithfulness constraints:

(16) *Word recognition*

<table>
<tr><th>/.paɾ.te./</th><th>*‘(nonsense)’</th><th>*|m|
/p/</th><th>*|m|
/b/</th><th>*|b|
/p/</th><th>*|p|
/b/</th></tr>
<tr><td>☞ |paɾte| ‘part’</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>|baɾte| ‘(nonsense)’</td><td>*!</td><td></td><td></td><td>*</td><td></td></tr>
<tr><td>|maɾte| ‘Mars’</td><td></td><td>*!</td><td></td><td></td><td></td></tr>
</table>

The winning candidate violates no constraints at all. A more interesting form is /.baɾ.te./:

(17) *Word recognition*

<table>
<tr><th>/.baɾ.te./</th><th>*‘(nonsense)’</th><th>*|m|
/p/</th><th>*|m|
/b/</th><th>*|b|
/p/</th><th>*|p|
/b/</th></tr>
<tr><td>☞ |paɾte| ‘part’</td><td></td><td></td><td></td><td></td><td>*</td></tr>
<tr><td>|baɾte| ‘(nonsense)’</td><td>*!</td><td></td><td></td><td></td><td></td></tr>
<tr><td>|maɾte| ‘March’</td><td></td><td></td><td>*!</td><td></td><td></td></tr>
</table>

In this case one still recognizes |paɾte| ‘part’, although a different ranking of some faithfulness constraints would have led one to recognize |maɾte| ‘Mars’ instead:

(18) *Word recognition*

<table>
<tr><th>/.baɾ.te./</th><th>*‘(nonsense)’</th><th>*|m|
/p/</th><th>*|p|
/b/</th><th>*|m|
/b/</th><th>*|b|
/p/</th></tr>
<tr><td>|paɾte| ‘part’</td><td></td><td></td><td>*!</td><td></td><td></td></tr>
<tr><td>|baɾte| ‘(nonsense)’</td><td>*!</td><td></td><td></td><td></td><td></td></tr>
<tr><td>☞ |maɾte| ‘Mars’</td><td></td><td></td><td></td><td>*</td><td></td></tr>
</table>

We cannot predict which of the two options people will choose. In any case, the choice between tableaus (17) and (18) does not depend on the degree of ambiguity of the auditory VOT.

5.2. The parallel edition of the process of phonetic-phonological comprehension

The situation is different in the parallel model of Figure 12.

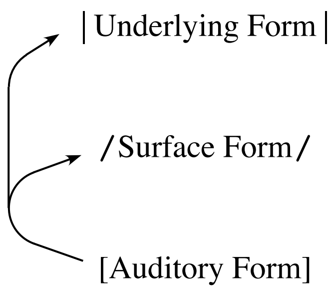

Fig. 12 The phonetic-phonological comprehension process, parallel edition.

We first provide a ranking that makes the listener perceive a VOT of –100 ms as /.baɾ.te./, never mind that the faithful lexical item |baɾte| does not exist. If the lexicon is still capable of telling the listener that the word the speaker intended was |paɾte|, the ranking can be that in tableau (19).

(19) *Perception possibly but not really influenced by lexical access*

[–100 ms]	*/p/ [–100]	*/b/ [+30]	*‘(non-sense)’	*|b| /p/	*|p| /b/	*/p/ [–20]	*/b/ [–20]	*/p/ [+30]	*/b/ [–100]
/.baɾ.te./ |baɾte|			*!						*
/.paɾ.te./ |baɾte|	*!		*	*					*
☞ /.baɾ.te./ |paɾte|					*				
/.paɾ.te./ |paɾte|	*!								

In the case of a VOT of –20 ms, which was perceived as /b/ in the sequential model, the perception now becomes /p/, as shown in tableau 20:

(20) *Perception possibly and really influenced by lexical access*

[–20 ms]	*/p/ [–100]	*/b/ [+30]	*‘(non-sense)’	*|b| /p/	*|p| /b/	*/p/ [–20]	*/b/ [–20]	*/p/ [+30]	*/b/ [–100]
/.baɾ.te./ |baɾte|			*!				*		
/.paɾ.te./ |baɾte|			*!	*		*	*		
/.baɾ.te./ |paɾte|					*!				
☞ /.paɾ.te./ |paɾte|						*			

In this tableau we see that the cue constraints prefer /b/, but the faithfulness constraint, forced by *‘(nonsense)’, prefers /p/.

A remaining question: can *‘(nonsense)’ ever be violated in a winning form? Yes, if it is outranked by faithfulness. In that case, tableau (19) would become tableau (21).

(21) *Recognizing a nonsense word*

[–100 ms]	*/p/ [–100]	*/b/ [+30]	*|b| /p/	*|p| /b/	*‘(non-sense)’	*/p/ [–20]	*/b/ [–20]	*/p/ [+30]	*/b/ [–100]
☞ /.baɾ.te./ |barte|					*				*
/.paɾ.te./ |barte|	*!		*		*				*
/.baɾ.te./ |parte|				*!					
/.paɾ.te./ |parte|	*!								

If both the cue constraints and the faithfulness constraints are ranked high enough, the auditory form is apparently capable of creating a new underlying form. You can see this explicitly in tableau (21), but you can also see it from Figure 1 if you regard the cue and faithfulness constraints in that figure as strong connections, and the lexical constraints as weak. Oops, now I have already talked about semantics...

6. The quadruplet Underlying – Surface – Auditory – Articulatory

6.1. The process of phonological-phonetic production

The typical process in this quadruplet is *phonological-phonetic production*. I think that I cannot model the phenomena that I want to model if I do not assume that this process is parallel, as in Figure 13. For instance, in **Boersma (2006a)**, faithfulness constraints must crucially interact with both articulatory and cue constraints.

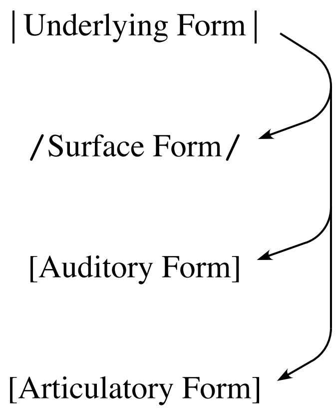

Fig. 13 The phonological-phonetic production process, fully parallel edition.

6.2. The acquisition of phonological-phonetic production

Boersma (2006a) shows that some observed universal rankings of faithfulness constraints (between Underlying Form and Surface Form) are predicted to be automatic results of parallel phonological-phonetic production:

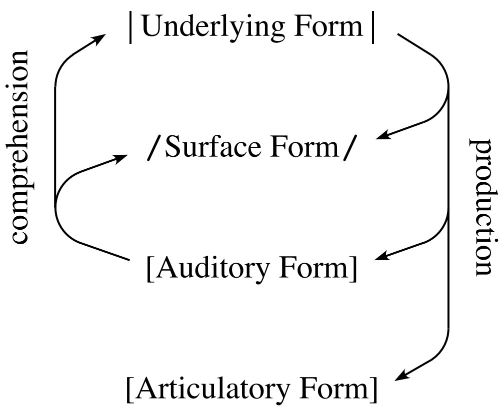

Fig. 14 Bidirectional acquisition of the phonological-phonetic production process.

The learning algorithm predicts rankings of faithfulness constraints by frequency and auditory cue quality, without the need for innately ranked positional faithfulness constraints (Beckman 1998), rankability by extralinguistic knowledge of auditory distances (Steriade's 2001 P-map), or rankability by linguistically computed confusability (Boersma 1998).

7. Semantic representations

Figure 1 includes only the two semantic representations that are of most interest to phonologists: the morpheme (for establishing morphemic identity), and the context (which influences expectations in comprehension). Semanticists would probably want to include more, such as the Semantic Form. They would also probably not regard the Morpheme as exclusively semantic, because e.g. the morpheme 'Nominative Singular' expresses a syntactic function rather than a semantic role. And semanticists have equivalent (or nearly equivalent) names for the Context (or Context Change), such as 'message meaning', 'situation', 'discourse representation structure', 'pragmatic context', or even 'pragmatic form', some of which suggest that this representation is not exclusively semantic either.

OT semanticists tend to be interested in the relation between Semantic Form and Context Change, or just between Form and Meaning, where 'Form' is the Morpheme (e.g. *him* or *himself*) and Meaning can be a part of the Context (e.g. the person referred to be the pronoun or anaphor). All this is far away from the interests of phonologists, but it is important to note that OT semanticists have invented *Bidirectional Optimality Theory* for the solution of their problems, especially for the problem of how to explain the difference between *to kill* and *to cause to die* (Blutner 2000).

8. The phonology-semantics interface: the lexicon

8.1. Relations

Since Saussure (1916), lexical entries have been regarded as 'form-meaning' pairs, whose 'form' part is the Underlying Form and whose 'meaning' part we can identify with the Morpheme. Saussure's own terms were *signifiant* ('signifier' = form) and *signifié* ('signified' = meaning), and he insisted that their relation is *arbitrary*, i.e. there are no cross-linguistic universals on what form goes with what meaning (except for some onomatopoeiae).

The relation between form and meaning in the lexicon is usually regarded as fixed. This happens even in OT. Boersma (1999/2001), for instance, used 'lexical constraints' such as *|ʀɑd|'wheel', but this example just expressed the listener's reluctance to access the single lexical item *|ʀɑd|'wheel', as opposed to, say, the item *|ʀɑt|'rat', with which it could be in competition during word recognition (in Dutch). Escudero (2005:214–236) went a bit further, proposing a competition between multiple lexical items with the same meaning, with 'lexical constraints' such as *|tʃika|'girl' and *|tʃɪka|'girl' for Dutch learners of Spanish. But only **Apoussidou (2006)** investigated the relation between Underlying Form and Morpheme as a violable **lexical constraint**, in an application to the interaction between lexical and grammatical stress in Greek. For instance, if the lexical constraint *|θalas|'sea' outranks *|θálas|'sea', then the morpheme 'sea' likes to be |θálas| (with lexically specified stress) in the Underlying Form, whereas if the ranking is the reverse, the morpheme 'sea' likes to be |θalas| (without any lexical stress specification) in the Underlying Form.

8.2. The process of lexical retrieval in production

If you ignore everything outside the lexicon (or if you have a serial modular view of production), then you will believe in the existence of a local process that retrieves the Underlying Form |dɔɡ|, given the Morpheme 'dog', as in Figure 15.

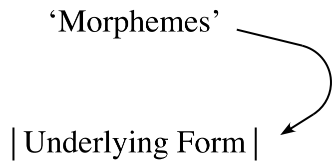

Fig. 15 The isolated lexical retrieval process in production.

8.3. The process of the access of meaning in comprehension

Analogously, if the listener has an Underlying Form at her disposal, she can access its meaning, as in Figure 16.

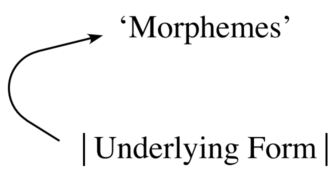

Fig. 16 The isolated lexical access process in comprehension.

8.4. The acquisition of lexical relations

The isolated acquisition of word-meaning pairs, as in Figure 17, is something that does not concern the phonologist much unless lower levels of representation are involved at the same time, as in §9.

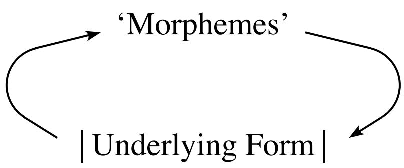

Fig. 17 Isolated bidirectional lexical acquisition.

9. The triplet Morphemes – Underlying Form – Surface Form

9.1. The influence of Morphemes (and Context) on word recognition

Boersma (1999/2001) modelled the connection from Surface Form to Underlying Form as an interaction of faithfulness constraints and lexical constraints, which militated against certain combinations of form and meaning. For instance, the mapping from the Dutch surface form /.ʀɑt./ to either the lexical form-meaning pair |ʀɑd|‘wheel’ or the lexical form-meaning pair |ʀɑt|‘rat’ would be decided by ‘lexical constraints’ that were conditioned by the Context, e.g. if the Context is “turn”, then *|ʀɑt|‘rat’/“turn” will probably outrank *|ʀɑd|‘wheel’/“turn”. In the present model, it would be the Morpheme-Context relation that decides this, i.e. the ranking *‘rat’“turn” >> *‘wheel’“turn”, but the idea is the same. As pointed out by Boersma (1999/2001), this solves Smolensky’s problem in (5) without having to invoke Hale & Reiss’ analysis-by-synthesis model. If we assume, just as an example, that |an| means ‘wheel’, |am| means ‘rat’, |aŋ| means ‘guinea pig’, and the context is “turn” (perhaps because |pa| means something like ‘turn’), tableau (5) can be corrected as tableau (22).

(22) *The success of phonological comprehension with semantics*

“turn” /.am.pa./	*CODA WITH SEPARATE PLACES$_S$	*‘guinea pig’ “turn”	*‘rat’ “turn”	*‘wheel’ “turn”	IDENT PLACE$_{US}$
☞ “turn” ‘wheel’ |an+pa| /.am.pa./				*	*
“turn” ‘rat’ |am+pa| /.am.pa./			*!		
“turn” ‘guinea pig’ |aŋ+pa| /.am.pa./		*!			*

As long as two of the three semantic (or reference) constraints outrank the faithfulness constraint for place, the listener will succeed in recognizing /.am.pa./ correctly, i.e. in finding the correct Underlying Form. For this to work, the semantic constraints have to be able to interact with the faithfulness constraints, i.e. word recognition and the access of meaning have to run in parallel, as in Figure 18.

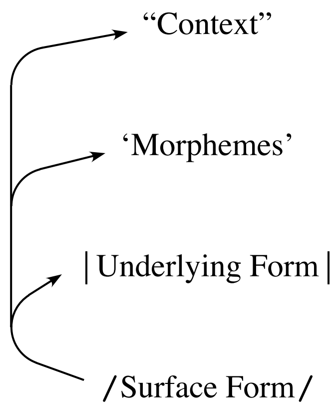

Fig. 18 Parallel access of lexical form and meaning.

9.2. Acquisition

Apoussidou (2006) shows that there is an automatic acquisition bias towards creating one Underlying Form for each morpheme, without the need for intelligent repair mechanisms like ‘surgery’ (Tesar, Alderete, Horwood, Merchant, Nishitani & Prince 2003):

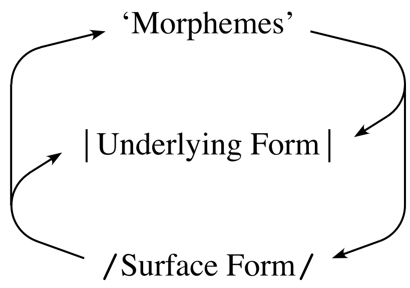

Fig. 19 The acquisition of underlying forms.

10. Whole-language simulations

In the above three triplet acquisition models I have been simplifying severely. A more realistic model of phonological-phonetic comprehension will include at least the quadruplet

Auditory Form – Surface Form – Underlying Form – Morphemes

where the Auditory Form and the Morphemes are known, but the Surface Form and the Underlying Form have to be *constructed* (in both senses of the word, i.e. gradually by the learner and on the fly by the listener). An example in the present paper are tableaus (19), (20), and (21), where a makeshift constraint *‘(nonsense)’ appears, that stands for a morphemic constraint. Because these tableaus have the Auditory Form as their input, they involve all four representations.

In production, the fifth representation, namely Articulatory Form, comes in as well. With these five we may start to be capable of doing **whole-language simulations**, i.e. computer simulations of the acquisition process that use as much realistic data from the language as possible (as well as transmission noise) and thereby derive the complete phonological-phonetic system of that language. Repeating the process for several generations should generate predictions about the stability and evolution of the sound system.

References (crucial material supplementing the present paper is given in bold)

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

Apoussidou, Diana, & Paul Boersma (2003). The learnability of Latin stress. *Proceedings of the Institute of Phonetic Sciences Amsterdam* **25**: 101–148.

Apoussidou, Diana, & Paul Boersma (2004). Comparing two Optimality-Theoretic learning algorithms for Latin stress. *WCCFL* **23**: 29–42.

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

Blutner, Reinhard (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics* **17**: 189–216.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings 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 dissertation, University of Amsterdam.

Boersma, Paul (2000). The OCP in the perception grammar. *Rutgers Optimality Archive* **435**.

Boersma, Paul (2001). Phonology-semantics interaction in OT, and its acquisition. In Robert Kirchner, Wolf Wikeley, & Joe Pater (eds.) *Papers in Experimental and Theoretical Linguistics*. Vol. 6. Edmonton: University of Alberta. 24–35. [*Rutgers Optimality Archive* **369**, 1999]

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

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

Boersma, Paul (2005b). Prototypicality judgments as inverted perception. Rutgers Optimality Archive **742**. To appear in a book edited by Gisbert Fanselow & Caroline Féry.

Boersma, Paul (2006a). The acquisition and evolution of faithfulness rankings. Talk at MFM 14, Manchester, May 27, 2006. [http://www.fon.hum.uva.nl/paul/]

Boersma, Paul (2006b). A constraint-based explanation of the McGurk effect. *Rutgers Optimality Archive* 869.

Boersma, Paul (2006c). The interaction of structural and cue constraints in phonological perception and production. Manuscript, University of Amsterdam.

Boersma, Paul, & Paola Escudero (2004). Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. *Rutgers Optimality Archive* **684**. [to appear in a book edited by Elan Dresher, Peter Avery, & Keren Rice]

Boersma, Paul, & Silke Hamann (2006). Sibilant inventories in bidirectional phonology and phonetics. Talk at OCP 3, Budapest, January 17, 2006. http://www.fon.hum.uva.nl/paul/

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

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

Johnson, Keith, Edward Flemming, and Richard Wright (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language* **69**: 505–528.

Kirchner, Robert (1998). *Lenition in phonetically-based Optimality Theory*. PhD dissertation, 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. pp. 249–384. [*Rutgers Optimality Archive* **60**]

McGurk, Harry, & John McDonald (1976). Hearing lips and seeing voices. *Nature* **264**: 746–748.

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.

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

Pater, Joe (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In René Kager, Joe Pater, & Wim Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press. 219–244.

Polivanov, Evgenij Dmitrievič (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* **4**: 79–96. [English translation: The subjective nature of the perceptions of language sounds. In E.D. Polivanov (1974): *Selected works: articles on general linguistics*. The Hague: Mouton. 223–237]

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

Smolensky, Paul (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry* **27**: 720–731.

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

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. [*Rutgers Optimality Archive* **177**]

Tesar, Bruce (1998). An iterative strategy for language learning. *Lingua* **104**: 131–145.

Tesar, Bruce (1999). Robust interpretive parsing in metrical stress theory. *WCCFL* **17**: 625–639. [*Rutgers Optimality Archive* **262**]

Tesar, Bruce, John Alderete, Graham Horwood, Nazaré Merchant, Koichi Nishitani & Alan Prince (2003). Surgery in language learning. *WCCFL* **22**: 477–490.

Tesar, Bruce, & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**: 229–268.

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