(to appear in University of Alberta Papers in Experimental and Theoretical Linguistics 6, ed. by Robert Kirchner, Wolf Wikeley, and Joe Pater)

Phonology-semantics interaction in OT, and its acquisition*

Paul Boersma, University of Amsterdam
December 18, 1999

Abstract

Smolensky's (1996) original proposal of a single grammar for production and comprehension does not seem to hold in the face of phonological alternations. If comprehension is to be modelled as an Optimality-Theoretic grammar, this recognition grammar must contain lexical-access constraints whose rankings depend on the semantic context and on frequency of occurrence. These lexical constraints interact with faithfulness constraints in much the same way as the structural constraints do in the production grammar. The lexical constraints come into existence by the violation of an anti-lexicalization constraint in the recognition grammar, and are subsequently automatically ranked in the appropriate confusion-minimizing order by a gradual learning algorithm.

1. Two maximally opposing grammar models

On one end of a spectrum, a grammar can be seen as a description of the possible utterances in a language. On the other end of this spectrum, grammars can be seen as descriptions of the processes of human speech production and comprehension. This paper is about the extent to which in a basically procedural view of phonology, the same grammar can be used for production and comprehension.

1.1 The procedural grammar model of functional phonology

Figure 1 is a graphical representation of the grammar model defended in Boersma (1998). It makes a principled distinction between the speaker and the listener. Thus, the "ART" constraints (against articulatory effort) are relevant only for production, and the "*LEX" constraints (against lexical access) are relevant only for comprehension.

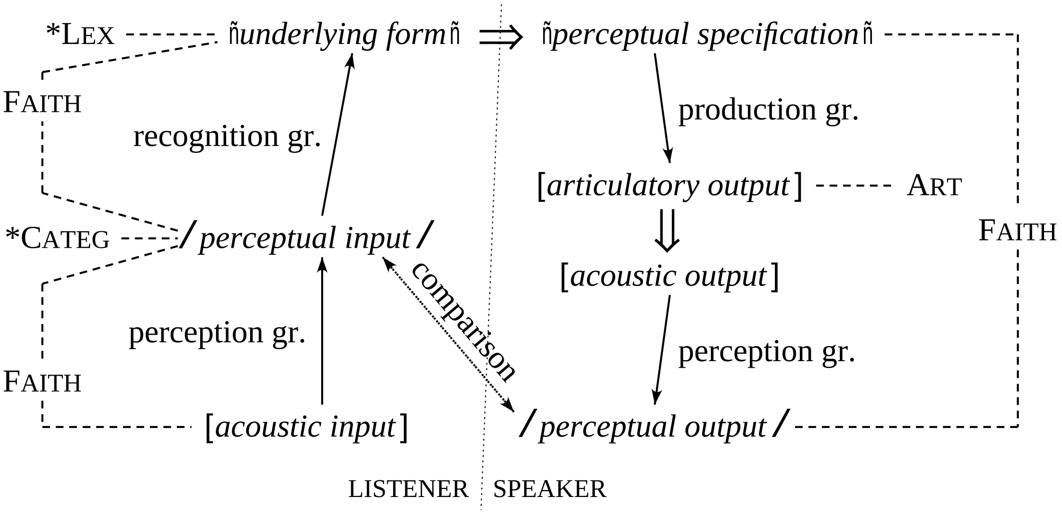


Fig. 1. The grammar model of functional phonology

1.2 The production grammar

As far as the speaker is concerned, functional phonology makes a distinction between articulation and perception: the production grammar consists of gestural constraints (ART), which implement the functional principle of minimation of articulatory effort, and of faithfulness constraints (“FAITH” on the right-hand side of Figure 1), which indirecty implement the functional principle of minimization of confusion.

1.3 The recognition grammar

The task of the recognition grammar (shown in the top left of Figure 1) is to map the perception of the utterance of another speaker to the underlying perceptual form, thus getting access to the lexicon and to meaning. Like the production grammar, it contains faithfulness constraints (“FAITH” in the top left of Figure 1). The recognition grammar also contains constraints against lexical access (*LEX). In this paper, I will show ***why*** these *LEX constraints must exist, and ***how*** the listener can learn them.

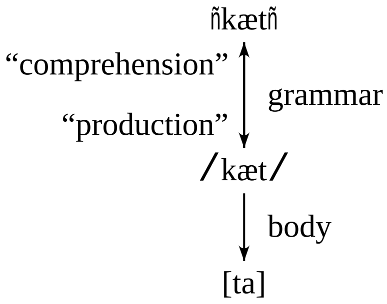
1.4 The perception grammar

The perception grammar, which appears twice in Figure 1, has many tasks in phonology (Boersma 1999). I will ignore it here, mentioning only that the two perceptual representations in Figure 1 (i.e. those written between slashes) are phonological surface forms, expressed in discrete perceptual features. What remains of Figure 1, then, is the speaker’s mapping of an underlying “perceptual specification” on a phonological surface form (production), and the listener’s mapping of a phonological surface form (as uttered by another speaker) on underlying lexical forms. In this view, there remain two sets of faithfulness constraints, which may or may not be shared between the two processing systems.

1.5 Structuralist grammar models

A common assumption in structuralist and generative grammar models is that the phonological part of the grammar should be used for production as well as comprehension. The grammar relates the underlying form with the phonological surface structure, and vice versa:

(2) *The structuralist grammar model (Hale & Reiss 1998)*



Hale & Reiss’s idea is that a child that says [ta] for the English word *cat*, must have a grammar output of /kæt/ in order to be able to map the adult utterance [kæt] to her underlying form |kæt|. This separation of grammar and body (or phonology and phonetics) still seems to be the prevailing grammar model nowadays.

2. Why and how performance and comprehension are modelled in OT

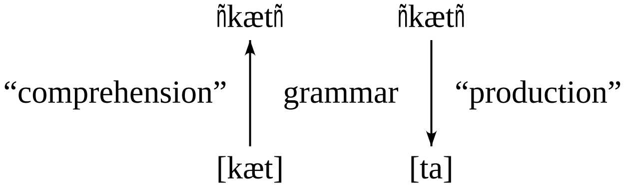
The literature on phonological acquisition (e.g. Gnanadesikan 1995, Smolensky 1996) is justified in using OT for describing performance, because the performance system of the learner shows systematic behaviour, *as if it belonged to the grammar*. For instance, the child's production of an underlying form |kæt| as [ta] can be described by an OT grammar in which some structural constraints outrank some faithfulness constraints:

(3) *Unfaithful production*

|kæt|	*CODA	PARSE
[kæt]	*!	
☞ [ta]		*

The generative grammar model by Smolensky (1996) assumes a monostratal relation between the underlying form and the phonetic form, unlike Hale & Reiss's. In his model, production and comprehension are handled by a single grammar, which, though it may map underlying |kæt| to surface [ta], should still be capable of comprehending an adult form [kæt] as |kæt|:

(4) *A single grammar for production and comprehension*



Here is how the same grammar as in (3) handles comprehension (Smolensky 1996: 725):

(5) *Faithful comprehension*

[kæt]	*CODA	PARSE
☞ |kæt|	*	
|skæti|	*	*!*

The top left cell now contains a surface form, and the candidates are underlying forms. The crucial insight (which invalidates Hale & Reiss's idea of grammar–body separation) is that *CODA is violated in every candidate, because this constraint evaluates the adult surface form [kæt], which is identical for both lexical candidates |kæt| and |skæti|. In comprehension, therefore, structural constraints do not contribute to determining the winning candidate. The winner must be determined exclusively by the faithfulness constraints. The maximally faithful form |kæt| will win, since it violates no faithfulness constraints at all. Smolensky's conclusion is that a single grammar can perform "production" as well as "comprehension".

3. When Smolensky's model does not work

As Hale and Reiss (1998: 661) pointed out, the fact that in Smolensky's comprehension model the most faithful candidate will always win, is problematic. Since GEN would always make available a candidate identical to the perceived form, comprehension would reduce to an identity mapping, which we would hardly have to model with an Optimality-Theoretic grammar. And since Smolensky's procedure cannot map an adult form to a lexical form that is different from it, this procedure is incapable of undoing phonological alternations.

As an example close to that of Hale & Reiss (1998: 661), I will consider final devoicing in Dutch, which causes the two words |rɑd| 'wheel' and |rɑt| 'rat' to merge on the surface (I assume here that the speaker has created an underlying ñdñ – ñtñ contrast on the basis of e.g. /raˇd´r´/ 'wheels' and /rAt´/ 'rats'). Denoting the structural constraint that causes final devoicing simply as *VOICEDCODA, and using square brackets for the surface forms, the unfaithful surfacing of |rɑd| 'wheel' is described by (6).

(6) *Unfaithful production of the wheel*

|rɑd| 'wheel'	*VOICEDCODA	MAXVOI
[rɑd]	*!	
☞ [rɑt]		*

The problem, now, is that this surface form [rɑt] will always be recognized as |rɑt| 'rat', even if the speaker obviously meant to refer to a wheel:

(7) *Failure to recognize the wheel*

[rɑt]	*VOICEDCODA	MAXVOI
☞ |rɑt| 'rat'		
|rɑd| 'wheel'		*

As before, we must note that the structural constraint *VOICEDCODA is not violated for any candidate.

Our conclusion must be that |rɑd| 'wheel' cannot be recognized, so phonological alternations cannot be handled by Smolensky's comprehension model. Looking into the literature on this model, we see that it has been tried on lexicalization of morphologically alternating pairs (Tesar & Smolensky 1996), prosodic parsing of surface stress sequences (Tesar 1996, 1997, to appear), and intermediate empty traces in syntax (Smolensky 1996), none of which will help in the comprehension of a single word form.

The source of the problem is that faithfulness violations are problematic in a faithfulness-only mapping: we lack constraints that directly evaluate the candidate lexical forms in comprehension. This is different from the production grammar, where structural constraints directly evaluate the candidate output forms.

4. The solution: lexical-access constraints

So we need constraints that directly evaluate the candidate lexical forms in the recognition grammar, analogous to the structural constraints in the production grammar, which directly evaluate surface forms. Fig. 1 shows these *Lex constraints, which ensure that every underlying form violates at least one constraint, so that the listener can choose on other criteria than just faithfulness. *Lex takes the form of a set of constraints against recognizing each lexical item:

(8) *Lex (*x*)
“do not recognize an utterance as the lexical item *x*”

Examples are *Lex (|rɑt| ‘rat’) and *Lex (|rɑd| ‘wheel’).

5. The dependence of *Lex on frequency

Psycholinguistic evidence shows that if everything else is equal, words with higher frequency are recognized better than words with lower frequency. Now, Dutch |rɑt| ‘rat’ is more common than |rɑd| ‘wheel’, because the normal word for ‘wheel’ is |ʋil|. So we expect that out of context, an utterance [rɑt] will tend to be recognized as |rɑt| ‘rat’. We can formalize this as a frequency-dependent ranking of *Lex constraints:

(9) *Dependence of *Lex on frequency*
*Lex (|rɑd| ‘wheel’) >> *Lex (|rɑt| ‘rat’)

The following tableau shows that [rɑt] will be recognized as |rɑt| ‘rat’, independently of the ranking of the faithfulness constraint MaxVoi:

(10) *A strong tendency to recognize the rat*

[rɑt]	*Lex (|rɑd| ‘wheel’)	*Voiced Coda	MaxVoi	*Lex (|rɑt| ‘rat’)
☞ |rɑt| ‘rat’				*
|rɑd| ‘wheel’	*!		*	

5.1 Learning the frequency dependence

The dependence of the ranking of *Lex constraints on frequency decreases the probability of misunderstanding. If [rɑt] means ‘rat’ 70% of the time, and ‘wheel’ 30% of the time, a listener who always recognizes ‘wheel’ will misunderstand the speaker 70% of the time, but a listener that recognizes ‘rat’ will misunderstand only 30% of the time. Since minimizing confusion is evolutionary advantageous (e.g. it maximizes knowledge of whether life-threatening danger comes from rodents or from automobiles), humans have probably an innate device that causes *Lex constraints to be higher ranked for less common words.

This innate device, I propose, is not something that has been created specifically for lexical access. Rather, it is the same independently needed **gradual learning algorithm** that does so many more good to the language learner (Boersma 1997; Boersma 1998: chs. 14-15; Boersma & Hayes 1999). This algorithm has the following ingredients:

(11) *Gradual learning algorithm*

(a) Each constraint has a ranking value along a continuous scale (ranking is not ordinal).
(b) Some noise is temporarily added to the ranking of each constraint at evaluation time.
(c) If the learner notices that the adult form differs from her own form, the learner will change her grammar
  (i) by raising the constraints violated in her own form
  (ii) and lowering the constraints violated in the adult form.

General properties of this algorithm are that it is convergent, gradual (it generates realistic learning curves: Boersma & Levelt 1999), and robust against variation and errors in the data. I will show how this algorithm leads to the frequency dependence in the *Lex family.

First suppose that the learner has a non-adult ranking, i.e., while [rɑt] should be mapped to |rɑt| 'rat' 70% of the time and to |rɑd| 'wheel' 30% of the time, the learner nevertheless has the ranking *Lex (|rɑt| 'rat') >> *Lex (|rɑd| 'wheel'), which leads to her recognizing the less common word |rɑd| 'wheel' all of the time (if we ignore the phonology for a moment).

Now we can distinguish two cases. The first case applies 30% of the time: it is when the adult means |rɑd| 'wheel', and the learner, of course, will recognize this correctly. According to the condition of error-drivenness (11c), the learner's grammar will not change. The second case, however, will apply 70% of the time: it is when the adult means |rɑt| 'rat', but the child comprehends |rɑd| 'wheel'. If the child notices the discrepancy, she will take a learning step:

(12) *Learning as a result of misrecognition*

[rɑt] from |rɑt| 'rat'	*Lex (|rɑt| 'rat')	*Lex (|rɑd| 'wheel')
√ |rɑt| 'rat'	*!→	
☞ |rɑd| 'wheel'		←*

Tableau (12) is to be interpreted as follows. The learner hears [rɑt], which is written in the top left cell because it becomes the input to the recognition grammar. The fact, noted by the learner, that this utterance should have been recognized as |rɑt| 'rat', is also written in the top left cell, so that this cell now contains all information available to the learner. There are now at least two relevant candidates for recognition, namely |rɑt| 'rat' and |rɑd| 'wheel'. The form |rɑd| 'wheel' is the winner in the learner's grammar (as we can see from the constraint violations), so it receives the traditional pointing finger. The candidate |rɑt| 'rat', however, is identified by the learner as being identical to the adult intended underlying form, and receives a check mark because the learner assumes it is correct; this will also mean that the learner must assume that her own winner is incorrect, which we depict by adding a couple of asterisks around the pointing finger. The learner will now take the two actions associated with a learning step: according to (11ci) she will raise the ranking of the constraint violated in her own form, namely *Lex (|rɑd| 'wheel'), which is shown as the leftward arrow in (12), and according to (11cii) she will lower the ranking of *Lex (|rɑt| 'rat'), which is shown by the rightward arrow.

5.2 The result of learning the frequency dependence: probability matching

During acquisition, *LEX (|rɑt| ‘rat’) will fall down *LEX (|rɑd| ‘wheel’), but not by much. The probability-matching property of the gradual learning algorithm, caused by the presence of the evaluation noise, will lead to an equilibrium in which the child will randomly choose |rɑt| ‘rat’ 70% of the time, and |rɑd| ‘wheel’ 30% of the time. Since this means that the child will still make an error 42% of the time (70%×30% + 30%×70%), some source of minimization of confusion must still be missing. That source, of course, is the semantic context.

6. The dependence of *LEX on semantic context

If the semantic context is ‘turn’, the recognition of ‘wheel’ is favoured over that of ‘rat’, as far as the lexicon is concerned. This can be formalized as a dependence of the ranking of the *LEX constraints on the semantic context:

(13) *Semantic conditioning of lexical access*

*LEX (|rɑt| ‘rat’ / context = ‘turn’) >> *LEX (|rɑd| ‘wheel’ / context = ‘turn’)

This will finally allow us to recognize |rɑd| ‘wheel’, even if both the phonology and the frequency would advise otherwise:

(14) *Recognizing the wheel*

[rɑt] context = ‘turn’	*LEX (|rɑt| ‘rat’ / ‘turn’)	*VOICED CODA	MAXVOI	*LEX (|rɑd| ‘wheel’ / ‘turn’)
|rɑt| ‘rat’	*!			
☞ |rɑd| ‘wheel’			*	*

6.1 Learning the semantic dependence

As with the frequency dependence, the evolutionary advantage of taking into account the semantic context is obvious. Another similarity is that we already have the innate device that leads to this confusion-minimizing behaviour, namely the same gradual learning algorithm.

Suppose, for instance, that the learner has a non-adult ranking, i.e., while a context of ‘turn’ should make her recognize [rɑt] as |rɑd| ‘wheel’, she nevertheless has the ranking *LEX (|rɑd| ‘wheel’ / ‘turn’) >> *LEX (|rɑt| ‘rat’ / ‘turn’), which causes her to recognize |rɑt| ‘rat’ all of the time. When the adult now says [rɑt] in the context ‘turn’, and means |rɑd| ‘wheel’, the child wrongly comprehends |rɑt| ‘rat’, and if she notices the discrepancy, she will *learn* from it:

(15) *Learning as a result of misrecognition*

[rɑt] context = ‘turn’ from |rɑd| ‘wheel’	*LEX (|rɑd| ‘wheel’ / ‘turn’)	*LEX (|rɑt| ‘rat’ / ‘turn’)
☞ |rɑt| ‘rat’		←*
√ |rɑd| ‘wheel’	*!→	

Eventually, these constraints will become ranked in the adult, functionally appropriate, order.

6.2 Connectionist interpretation of the semantic dependence model

In a simple neural-net analogy, each of the (say) one thousand possible atomic semantic contexts is represented by a neuron that says whether this context applies in the current communicative situation, and each of the (say) ten thousand lexical entries (lexical entry = Saussurean sign = phonological form + semantic concept) is represented by a neuron that is connected with all atomic semantic contexts. Thus these 11 thousand neurons are wired with ten million synaptic connections, each with its own strength. The learning step in (15) is then understood as follows:

(16) *Connectionist analogue of OT semantics-dependent comprehension*

(a) lower the synaptic strength between the context ‘turn’ and the sign |rɑt| ‘rat’;
(b) raise the synaptic strength between the context ‘turn’ and the sign |rɑd| ‘wheel’.

Thus, 11 thousand neurons implement ten million constraints. The frequency dependency of the *LEX constraints corresponds to generally stronger synaptic weights between contexts and common lexical entries.

The simple analogy breaks down in cases where two (or more) semantic contexts apply. Since synaptic strengths contribute additively to postsynaptic potentials, the two contexts must contribute subtractively to the ranking of *LEX (|rɑt| ‘rat’), thus generating perhaps 10 milliard *LEX constraints with double semantic conditions; these cannot be OT-rankable constraints, since their rankings must depend on the rankings of their constituents.

7. Interaction of *LEX with faithfulness

In (14), comprehension of ‘wheel’ only works if voicing faithfulness is outranked, i.e. if the semantics dominates the phonology. To see that there is a genuine factorial interaction between the phonology and the semantics, we have a third candidate |ʋil| ‘wheel’ come on stage. This candidate is a much more common lexical item in Dutch than |rɑd| ‘wheel’, so we expect the following ranking:

(17) *Ranking by frequency as well as by semantic context*

*LEX (|rɑt| ‘rat’ / context = ‘turn’)
\>> *LEX (|rɑd| ‘wheel’ / context = ‘turn’) (rank by context)
\>> *LEX (|ʋil| ‘wheel’ / context = ‘turn’) (rank by frequency)

If it were for this semantic ranking alone, [rɑt] would be recognized as |ʋil| ‘wheel’, which feels absurd phonologically. In this case, therefore, the phonology must dominate some of the semantics:

(18) *True phonology-semantics interaction*

[rɑt] context = ‘turn’	*LEX (|rɑt| ‘rat’ / ‘turn’)	*VOICED CODA	FAITH	*LEX (|rɑd| ‘wheel’ / ‘turn’)	*LEX (|ʋil| ‘wheel’ / ‘turn’)
|rɑt| ‘rat’	*!				
☞ |rɑd| ‘wheel’			*	*	
|ʋil| ‘wheel’			**!		*

We see that candidate 1 loses because it is semantically too distant from ‘turn’, and that candidate 3 loses because it is phonologically too distant from [rɑt]. In order to make |rɑd| ‘wheel’ the winner, faithfulness must be crucially sandwiched between two semantic constraints: *LEX (|rɑt| ‘rat’ / ‘turn’) >> FAITH >> *LEX (|rɑd| ‘wheel’ / ‘turn’). This result cannot be achieved in a grammar model in which the semantics follows the phonology; rather, it suggests that phonology and semantics interact in parallel.

7.1 Learning the phonology-semantics interaction

As in §5.1 and §6.1, we again suppose that a certain learner has a non-adult ranking, in this case an underdeveloped sensitivity to semantic context or an excessive sensitivity to phonological faithfulness: FAITH >> *LEX (|rɑt| ‘rat’ / ‘turn’). If the adult says [rɑt] in the context ‘turn’, and means |rɑd| ‘wheel’, the child will wrongly comprehend |rɑt| ‘rat’. If she notices the discrepancy, she will take a learning step:

(19) *Learning the phonology-semantics interaction*

[rɑt] context = ‘turn’ from |rɑd| ‘wheel’	FAITH	*LEX (|rɑt| ‘rat’ / ‘turn’)	*LEX (|rɑd| ‘wheel’ / ‘turn’)
☞ |rɑt| ‘rat’		←∗	
√ |rɑd| ‘wheel’	∗!→		∗→

This will eventually contribute to successful low-confusion communication. Note that as a side effect, *LEX (|rɑd| ‘wheel’ / ‘turn’) is demoted, though its violation in the adult form has nothing to do with the miscomprehension; the learning algorithm is still blind.

7.2 A more accurate view: distinguish separate faithfulness constraints

As Dutch codas often change their voicing, and Dutch vowels do not change their heights, Dutch coda-voicing faithfulness (MAXVOI) is probably ranked low, and vowel-height faithfulness (MAXHI) will be ranked high, even in the recognition grammar, so a more accurate account of the facts would have to make a distinction between the various faithfulness constraints:

(20) *Distinguishing the various faithfulness constraints*

[rɑt] context = ‘turn’	*LEX (|rɑt| ‘rat’ / ‘turn’)	*VOICED CODA	MAX HI	*LEX (|rɑd| ‘wheel’ / ‘turn’)	*LEX (|ʋil| ‘wheel’ / ‘turn’)	MAX VOI
|rɑt| ‘rat’	∗!					
☞ |rɑd| ‘wheel’				∗		∗
|ʋil| ‘wheel’			∗!		∗	∗

This captures the intuition that if we hear [rɑt], an underlying form |ʋil| would be phonologically too far off, but an underlying form |rɑd| would not.

8. Combining the grammars

8.1 Can the production and comprehension procedures share their Faith constraints?

The relative ranking of the faithfulness constraints MaxVoi and MaxHi in the recognition grammar proposed in (20) may well be the same as the ranking of their counterparts in the production grammar. After all, underlying forms like |ʋil| will never be pronounced [rɑt], whereas underlying forms like |rɑd| will. Surely MaxHi >> MaxVoi in production as well as in comprehension.

This symmetry between the two grammars suggests that they might share the faithfulness constraints and their rankings, i.e. that it is equally bad for the speaker to implement |*x*| as [*y*], as it is for the listener to recognize [*y*] as |*x*|. We may note that even functional phonology (Figure 1) could allow this sharing, since all representations involved are of a perceptual nature (the tetragon |underlying| - |spec| - /output/ - /input/).

8.2 Why would it be good to share Faith constraints?

If the recognition and production grammars shared their faithfulness constraints, this would implement the reciprocity of the Saussurean sign (Saussure 1916: 99). Hurford (1989) showed that this strategy has evolutionary advantages.

8.3 Can production and comprehension be handled by a single grammar after all?

If we take the bold move of including the lexical-access constraints in the *production grammar* as well, we see that these constraints would apply vacuously in that grammar:

(21) *Vacuous behaviour of *Lex in the production grammar*

|rɑd| ‘wheel’	*Lex (|rɑd|)	*VoicedCoda	Faith	*Lex (|rɑt|)
[rɑd]	*	*!		
☞ [rɑt]	*		*	

The underlying form is the same for all candidates, so every candidate violates *Lex (|rɑd|). We see here an analogy between the production grammar and the recognition grammar: Art works vacuously in comprehension, *Lex works vacuously in production. So it is possible to regard the speaker/listener as having a single grammar containing all three families of constraints:

(22) *Contents of a combined grammar for production and comprehension*

a. structural constraints, which are only discriminative during production;
b. lexical-access constraints, which are only discriminative during comprehension;
c. faithfulness constraints, which work for production as well as comprehension.

9. Bootstrapping: creation of *Lex constraints

If the perceived form is too different from any lexical form in the given semantic context, the listener should create a new lexical entry. This can be modelled by a *Lexicalize constraint. If the word |rɑd| ‘wheel’ did not yet exist in the lexicon, it would be created:

(23) *The creation of a new lexical entry*

[rɑt]	*Lex (|rɑt| ‘rat’ / ‘turn’)	*Voiced Coda	MaxHi	*Lexicalize	*Lex (|ʋil| ‘wheel’ / ‘turn’)	MaxVoi
|rɑt| ‘rat’	*!					
|ʋil| ‘wheel’			*!		*	
☞ |rɑt| ‘wheel’				*		

Note that the new lexical item is created faithfully, since the violations of any other candidate new item would form a superset of those of |rɑt| ‘wheel’. The appropriate underlying voicing of the final consonant must be learned later on the basis of morphological alternations.

With the creation of the new lexical entry |rɑt| ‘wheel’, the learner will also create a family of *Lex (|rɑt| ‘wheel’) constraints, which she will subsequently rank automatically on the basis of frequency and context.

10. Counterexample: a counterfeeding relation in Catalan

Consider the following data on final nasals in Catalan, with rather abstract underlying forms:

(24) *Catalan nasal finals*

|bint| ‘twenty’ → /bin/	cf. /bint-i-siŋ/ ‘twenty-five’
|bin| ‘wine’ → /bi/	cf. /bin-s/ ‘wines’

This chain shift can be handled in an OT production grammar by using local conjunction of faithfulness constraints, i.e., it is worse to delete both |n| and |t| than to delete only one of them:

(25) *Successful production of ‘twenty’ in Catalan*

|bint|	*Delete (nt / _#)	*[nt] / _#	*[n] / _#	*Delete (t / n_#)	*Delete (n / _#)
[bint]		*!			
☞ [bin]			*	*	
[bi]	*!				

(26) *Successful production of wine in Catalunya*

|bin|	*Delete (nt / _#)	*[nt] / _#	*[n] / _#	*Delete (t / n_#)	*Delete (n / _#)
[bin]			*!		
☞ [bi]					*

But comprehension still cannot be handled, since [bin] must never be recognized as |bin| ‘wine’, not even if the context favours it:

(27) *Failure to recognize ‘twenty’ in Catalan*

[bin] context = ‘drink’ from |bint| ‘twenty’	*Delete (nt / _#)	*[nt] / _#	*[n] / _#	*Delete (t / n_#)	*Delete (n / _#)	*Lex (|bint| ‘twenty’ / ‘drink’)	*Lex (|bin| ‘wine’ / ‘drink’)
☞ |bin| ‘wine’			*				←*
√ |bint| ‘twenty’			*	*!→		*→	

This kind of learning will eventually lead to a perverse ranking of the lexical-access constraints in the context ‘drink’. The cause of this problem is that there is a superset relation between the faithfulness constraints. Thus, the recognition grammar is more sensitive to opaque relations than the production grammar. To solve the problem, we may have to decide that underlying forms like |bint| ‘twenty’ and |bin| ‘wine’ are too abstract for real speakers, or else take recourse to a strategy like *sympathy* (McCarthy 1998), for which nobody has yet devised a learning algorithm.

11. Conclusion

Phonology and semantics must be handled in parallel in an OT recognition grammar containing constraints against lexical access. Smolensky’s (1996) idea of handling production and comprehension in a single grammar may well survive this extension.

Notes

* The author can be reached at paul.boersma@hum.uva.nl.

References

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

Boersma, Paul (1998). *Functional phonology*. PhD dissertation, University of Amsterdam. The Hague: Holland Academic Graphics. [http://www.fon.hum.uva.nl/paul/]

Boersma, Paul (1999). *On the need for a separate perception grammar.* [Rutgers Optimality Archive ROA-358]

Boersma, Paul & Bruce Hayes (1999). *Empirical tests of the Gradual Learning Algorithm*. Ms. University of Amsterdam and UCLA. [Rutgers Optimality Archive ROA-348, http://ruccs.rutgers.edu/roa.html]

Boersma, Paul & Clara Levelt (1999). Gradual constraint-ranking learning algorithm predicts acquisition order. To appear in *Proceedings of Child Language Research Forum* **30**. Stanford, Calif.: CSLI. [ROA-361]

Hale, Mark & Charles Reiss (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* 29: 656–683.

Hurford, James (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* 77: 187–222.

McCarthy, John J. (1998). *Sympathy and phonological opacity*. Ms. Univ. of Massachusetts, Amherst. [ROA-252]

Saussure, Ferdinand de (1916). *Cours de linguistique générale*. Edited by Charles Bally & Albert Sechehaye in collaboration with Albert Riedlinger. Paris: Payot & C^{ie}. [2nd edition, 1922]

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

Tesar, Bruce (1996). An iterative strategy for learning metrical stress in Optimality Theory. To appear in *Proceedings of the 21st Annual Boston University Conference on Language Development*.

Tesar, Bruce (1997). An iterative strategy for language learning.” *Lingua* 104: 131–145.

Tesar, Bruce (to appear). On the roles of optimality and strict domination in language learning. *Optimality Theory: Phonology, syntax, and acquisition*, edited by Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer. Oxford University Press.

Tesar, Bruce & Paul Smolensky (1996): *Learnability in Optimality Theory (long version)*. Technical Report 96-3, Dept. of Cognitive Science, Johns Hopkins University, Baltimore. [Rutgers Optimality Archive ROA-156]