

Phonetically-driven acquisition of phonology

Paul Boersma

University of Amsterdam

March 31, 2000

In this paper, I will show how the substantive content of phonological constraints (i.e. any reference to specific phonological features or structures) is learned during the course of acquisition, i.e. how phonological development is driven by the language-specific development of articulation and perception.

1. Grammar model

In order to be able to state more precisely and formally how acquisition develops, we need an explicit model of the processes of speech production and comprehension. With the theory of functional phonology (Boersma 1998), I proposed that these processes can be described with three Optimality-Theoretic grammars (Figure 1).

First, the **production grammar** maps an underlying form, expressed in perceptual specifications, to a continuous articulatory output form, which is then converted by the speaker's perception system to a more discrete perceptual output form. More formally, the output is chosen from a list of relevant output candidates, and the winning candidate is the one that minimally violates the ranked constraints of the production grammar. The constraints are divided into two groups: articulatory constraints (ART), which evaluate each articulatory output, thus implementing the functional principle of minimizing articulatory effort, and faithfulness constraints (FAITH), which evaluate the similarity between each perceptual output and the underlying form, thus implementing the functional principle of minimizing perceptual confusion. This order of processing (production followed by perception, with a comparison between the perceptual result and its specification) is a typical

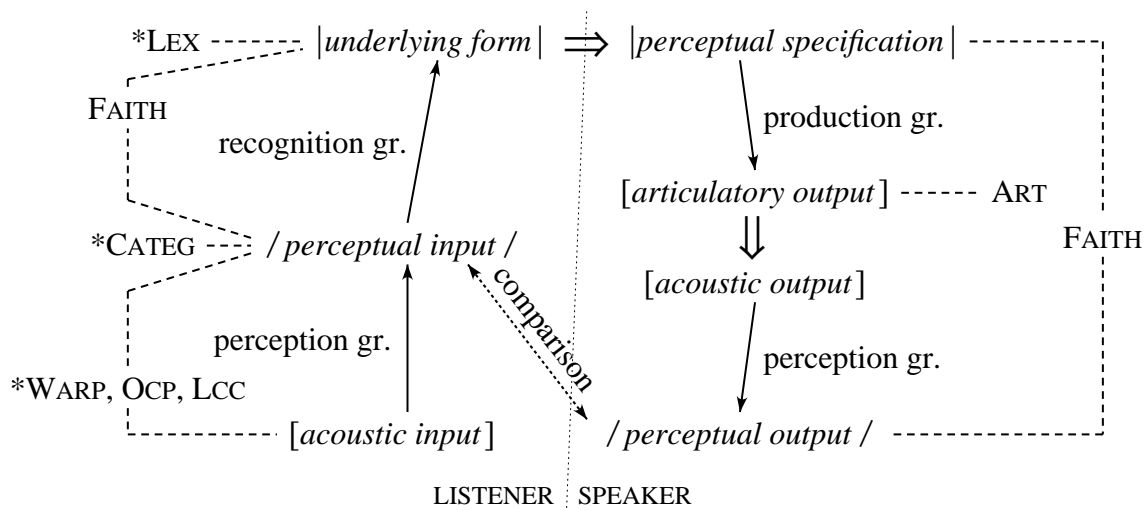


Fig. 1 The grammar model of functional phonology

example of the general perceptual control loop (Powers 1973), and shows that speech production is perception-oriented, like all human behaviour.

Secondly, the **perception grammar** consists of constraints that help to classify the acoustic input to the ear into a finite number of perceptual categories (*CATEG, *WARP) and higher-level structures (OCP, LCC). This grammar was mentioned above as a part of the speech production process, but is also used by the listener as a first step in the comprehension process.

Thirdly, the **recognition grammar** maps the discrete output of the perception grammar to underlying lexical forms. It consists of constraints that evaluate the lexical and semantic appropriateness of recognized underlying forms (*LEX) and, as in the production grammar, faithfulness constraints (FAITH).

A comprehensive acquisition model has to account for the development of all these three grammars.

2. The adult perception grammar

The task of the perception grammar (Boersma 1998, 1999a) is to abstract raw acoustic material, with its dependence on the age, sex, physiology, and state of the speaker, on room acoustics and the weather, and on some more random-like causes of variation, into a more reproducible (probably discrete) representation that is more suited for lexical access and that can, for purposes of learning, be compared with the output of the listener's own production grammar. Two questions have to be answered. First, is there indeed a modular division between perception grammar and recognition grammar, as shown in the model of Figure 1? Secondly, what is the nature of the intermediate representation, which is called "perceptual input" in Figure 1? For both of these questions, we can try to find answers in the psycholinguistic literature.

2.1 Lexical effects in the perception grammar?

If the perception and recognition grammars are two sequentially ordered modules, and only the recognition grammar can take lexical information into account, then we expect to see no influence of lexical information in the processes of the perception grammar. Since one of the tasks of the perception grammar is to convert continuous acoustic feature values into discrete phonological perceptual feature values, we must expect that lexical information has no influence on this categorization. With Saussure (1916), I will view a lexical entry as a simple combination of a perceptual underlying form (*signifiant*) and a meaning (*signifié*). So we expect that there is no influence of the perceptual underlying form, so that, for instance, the existence of the English words |b̥i:f| 'beef' and |p^hi:s| 'peace' next to the non-existent *|p^hi:f| and *|b̥i:s| should not lead to a bias in the perception of the first sound in the ambiguous acoustic inputs [pi:f] and [pi:s] as /b̥/ and /p^h/, respectively. And we expect that there is no influence of the semantic context, so that, for instance, the ambiguous acoustic input [pæ:θ] should not lead to a perception bias for |b̥æ:θ| 'bath' in a 'hot water' context and for |p^hæ:θ| 'path' in a 'jogging' context. So we expect that intermediate acoustic forms with a voice onset time (VOT) below, say, 36 ms are perceived as /b̥/ (which has a typical VOT in

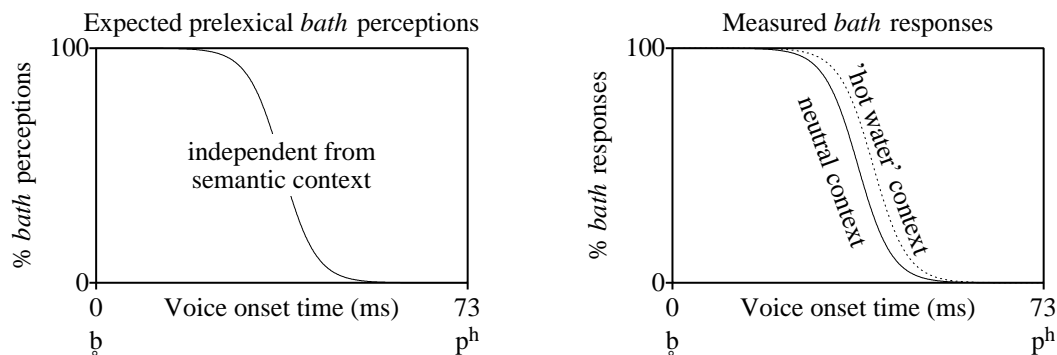


Fig. 2 Expected and measured bath/path category boundaries in the semantic context 'hot water'

English of 0 ms) and that those with a voice onset time above 36 ms will be perceived as /p^h/ (which has a typical VOT of 73 ms in English), independently from the semantic context. Figure 2 (left) shows this expected situation, with a transitional region around the category boundary of 36 ms.

But lexical shifts of category boundaries have been found repeatedly in psycholinguistic experiments (Bagley 1900, Garnes & Bond 1976, Ganong 1980), which seems at first glance to constitute a problem for our modular model of comprehension. Miller, Green & Schermer (1984) found that if listeners were required to attend to the sentence that provided the semantic context, they would shift the *bath/path* category boundary in a semantically appropriate way, by approximately 3 ms along the voicing continuum (Figure 2, right). This effect is small (about 4% of the English /b/-/p^h/ distance) but real, and requires an explanation. For instance, the subjects could have used several strategies for their responses, including basing their choice on a later established semantic context instead of on early prelexical perception alone. That is, if I ask someone to classify the acoustically ambiguous [pæ:θ], how do I know whether I am tapping the output of her perception grammar or the output of her recognition grammar (Figure 3)? Fortunately, there are ways to find out which of the grammars we are tapping, like measuring response times (larger latencies mean higher-level processing), manipulating the response speed (faster responses will be influenced less by late processing), promising rewards (which will have little effect on early processing), and manipulating conscious attention (which will mainly influence later processing). Thus, Miller et al., well aware of this possible distinction between perceived and reported categories (for an explicit discussion, see Samuel 1990), gave the listeners a second task, in which they were no longer required to consciously attend to the semantic context. In this speeded task, the lexical bias 'vanished',¹ which suggests that the influence of the lexicon does not occur in prelexical perception.

¹ Miller et al. made the classical mistake of interpreting the fact that they found no significant effect in the second task, as the absence of an effect. Fortunately, they performed all their experiments twice, so that the outcome may be marginally reliable after all: if we daringly reconstruct the standard errors from Miller et al.'s significance reports, we find that the difference between the 'hot water' context and the neutral context in the conscious-attention condition was between +1.4 ms and +5.0 ms (95% confidence interval), and that this difference in the second condition was between -1.3 ms and +2.3 ms (which is different from 'vanishing'). The difference between the two tasks, then, was between +0.2 and +5.4 ms (95% confidence), i.e. the influence of the semantic context was 'reliably', though perhaps not much, greater in the high-attention task than in the low-attention task.

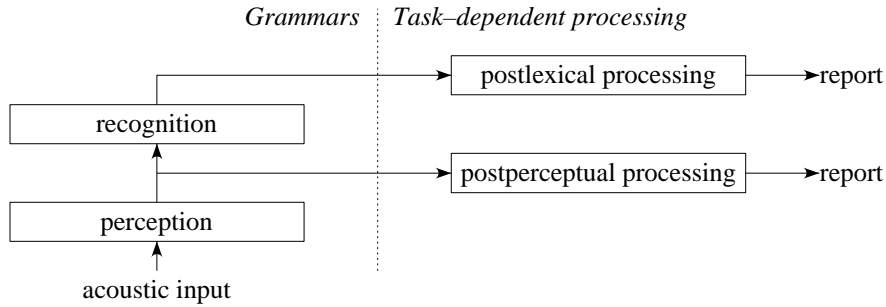


Fig. 3 Modular view of the phoneme identification task, compatible with Figure 1 and with most, perhaps all, psycholinguistic experiments

Similar findings as for the lexical choice bias were found for lexical status, i.e. whether or not the word occurs in the lexicon (Ganong 1980). For instance, Miller & Dexter (1988) found that listeners sometimes preferred *beef* and *peace* for acoustically ambiguous inputs.

The two types of lexical influences mentioned above occurred only if the listeners were given enough time to look up lexical items. When they were forced to answer quickly, they showed no lexical biases at all. So it seems as if fast responses come from the perception grammar, and slow responses from the recognition grammar, even for a task (phoneme identification) for which the perception grammar seems sufficient. But of course, the comprehension system is optimized for comprehension, not phoneme identification, so the recognition grammar cannot be turned off and will influence the outcome of many so-called perception tasks, simply because the respondents cannot separate the output of the perception grammar from the output of the recognition grammar. According to the ‘race model’ of speech comprehension (Cutler, Mehler, Norris & Segui 1987)², the response of the subject will depend on which of the two routes (the pre-lexical perception grammar or the combination of perception and lexical access) will give the fastest answer (Figure 3). In the case of word-initial consonants, the perception grammar will often be fastest. In the case of word-final consonants, as in McQueen’s (1991) **fiss/fish* versus *kiss/*kish* experiment, the recognition system will have had time to home in on a small set of candidates before the ambiguous sibilant fricative [ç] is sounded. As soon as the perception grammar reports a sibilant, the recognition grammar, having heard /fi/, knows that the word must be [fi] ‘fish’, while the perception grammar itself is still busy finding out whether this [ç] is nearer to the /s/ category or to the /ʃ/ category.

The evidence seems compatible with the bottom-up view of Figure 1, if we take into account that what listeners report that they hear, is not necessarily identical to the output of the perception grammar (Figure 3). We must note here that the modular view is not compatible with the TRACE model of comprehension (McClelland & Elman 1986, Elman & McClelland 1988, McClelland 1991), which features a top-down path for facilitation of phoneme identification based on lexical expectations. For a race-like account of phenomena that had seemed to point to TRACE-like top-down processing, see Norris’ (1993, 1994) Shortlist model.

² Unfortunately, these authors confuse statistical significance with effect size, as do most psycholinguists.

2.2 Continuous contents of the output of the perception grammar

While the phoneme shifts mentioned in §2.1 constitute no evidence against the modular view of comprehension shown in Figure 1, they do seem to require that the output of the perception grammar does not solely consist of discrete phonological structures. In all of the above-mentioned experiments, it was only the acoustically ambiguous inputs that could show lexical bias effects: an outright [b̥æ:θ] or [p^hæ:θ] was always identified as |b̥æ:θ| or |p^hæ:θ|, respectively. This means that if the recognition grammar has to work with the output of the perception grammar, this representation should contain information on the phonetic qualities of the plosives. Otherwise, the recognition grammar would just copy the discretized output of the perception grammar even in the ambiguous cases, and no semantic bias could result, or the semantic bias would occur in all cases, giving a horizontal line in Figure 2 (right) instead of a cumulative Gaussian. At least, this is the interpretation of Ganong (1980)³ for lexical status (“lexical status has an effect before acoustic information is replaced by a phonetic categorization”) and of Miller et al. for the lexical choice bias (“the context effect was not due to a late decision process that operated on the outcome of an earlier process responsible for the analysis/interpretation of the acoustic input as specifying /b/ and /p/”)⁴. However, the listener, hearing /p^hæ:θ/ but noticing that |b̥æ:θ| would be semantically more appropriate, could have her perception grammar re-evaluate the acoustic input, which is still lingering in sensory memory, another ten times, and see whether it produces the discrete form /b̥æ:θ/ in at least one case. If so, the acoustic form can be considered consistent with |b̥æ:θ|, and the listener will report having heard /b̥/. This account (which is one of the three explanations considered by Ganong 1980) would explain why the boundary shift in Figure 2 (right) stays within the region of variable perception. See Boersma & Hayes (1999: §5) for a similar frequency-based account of gradient well-formedness.

To sum up, there seems to be no compelling reason to question the simple modular model in Figure 1. For the time being, I will assume that the perception grammar usually produces discrete outputs (an exception is the perception of a sound acoustically far outside any category, see Boersma 1998: 166), and that the repetitive evaluation proposed above is not needed in the normal communicative situation, other than in the highly artificial tasks that are often used in psycholinguistic experiments.

2.3 Discrete contents of the output of the perception grammar

Most researchers agree that there is a stage of phonetic categorization into discrete units (a notable exception is Klatt 1980, whose model accesses the lexicon directly from a neural spectrogram). The output of the perception grammar (the “perceptual input” of Figure 1), therefore, contains discrete features and structures. Several proposals have been made as to the nature of these pre-lexical units of perception: they could be phonetic features (Eimas & Corbit 1973), phonological features (Lahiri & Marslen-Wilson 1991), allophones (Wickelgren 1969), phonemes (Foss & Blanck 1980; Norris & Cutler 1988), syllables (Foss

³ Who, unfortunately, chose to selectively throw away all the data that would have been most likely to falsify his hypothesis.

⁴ As many 20th-century anglophone researchers, Miller et al. use the orthography-based symbols “b” and “p” to denote the English lenis voiceless plosive and aspirated plosive, which they call ‘voiced’ and ‘voiceless’ (while the IPA voiced [b] is called ‘pre-voiced’).

& Swinney 1973; Cole & Scott 1974; Segui, Frauenfelder & Mehler 1981; Mehler, Dommergues, Frauenfelder & Segui 1981; Dupoux & Mehler 1990), or even articulatory gestures (Liberman & Mattingly 1985). The results of these experiments tend to be highly task-dependent. As Pisoni & Luce (1987) and McQueen & Cutler (1997) note, the listener probably constructs many of these units at the same time, and it is this parallel interpretation of perceptual abstraction that comes natural in an Optimality-Theoretic account of perception, as defended in Boersma (1999a: §7.7). In the following, therefore, I shall concentrate on the phonologically most interesting tasks of the perception grammar, namely categorization and sequential abstraction, and ignore other tasks (like rate normalization and speaker normalization).

2.4 Categorization and acoustics-to-perception faithfulness

Consider the perceptual feature of vowel height. Acoustically, this derives from the height of the first formant (F1) of the vowel. We assume that the perceptual dimension can also be expressed in terms of F1 values, so that we can compare the acoustic input to the perception grammar (a raw F1 value) with the output of the perception grammar, which is a more discrete perceptual representation. Suppose that the listener divides the perceptual feature of vowel height into four categories, which can be labelled /low/, /lower mid/, /higher mid/, and /high/. Along the perceptual F1 dimension, they can correspond to the categories /maximum F1/, /F1 = 600 Hz/, /F1 = 400 Hz/, and /minimum F1/, respectively. Let's limit the discussion to front vowels. The four categories can then be labelled /a/, /ε/, /e/, and /i/, respectively.

As a first approximation, we know that listeners classify an incoming acoustic feature value into the nearest perceptual category. Thus, a front vowel with an F1 of 450 Hz will probably be perceived as /e/ (F1 = 400 Hz), not as /ε/ (F1 = 600 Hz). According to Boersma (1998: §8.3), this situation can be described in constraint language as an interaction of anti-categorization constraints (*CATEG) and perceptual faithfulness constraints (*WARP):

(1) *CATEG (*f*: *y*)

“Do not perceive an output value (i.e. “perceptual input” in Figure 1) *y* on the perceptual tier *f*.”

For our example of four vowel heights, the constraints *CATEG (F1: 400 Hz) and *CATEG (F2: 600 Hz) are ranked relatively low, whereas *CATEG (F1: *y*) is ranked high for all *y* between 400 and 600 Hz.

The faithfulness constraints demand that acoustic inputs are mapped to perceptual inputs that are not far removed from them:

(2) *WARP (*f*: *x*, *y*)

“Do not perceive an acoustic input *x* on the perceptual tier *f* as a different value *y*.”

If *CATEG and *WARP were the only constraints in the perception grammar, the listener would be best off not perceiving an incoming F1 value at all, thereby vacuously satisfying all *CATEG and *WARP constraints. Therefore, the listener needs a perceptual correspondence

constraint to ensure that the output of the perception grammar contains something that corresponds to the input:

(3) PERCEIVE ($f: x$):

“If the acoustic input contains the value x on the perceptual tier f , the output of the perception grammar should contain a corresponding value.”


2.5 Functionally desirable ranking of categorization constraints

There are three desirable properties for ranking *WARP constraints, all of them based on considerations of minimization of confusion: the ranking should favour perception into a near category, it should favour perception into the nearest category, and it should favour perception into the most likely category.

First, the ranking of *WARP should depend on the one-sided perceptual distance between x and y . For instance, *WARP (F1: 450 Hz, 400 Hz) should be ranked higher than *WARP (F1: 420 Hz, 400 Hz), which means that it should be worse for the listener to perceive a front vowel with an F1 of 450 Hz as /e/ than it should be for her to perceive a front vowel with an F1 of 420 as /e/. Likewise, *WARP (F1: 300 Hz, 400 Hz) should outrank *WARP (F1: 300 Hz, 600 Hz): it is worse to perceive the pronunciation [i] as /e/ than as /ε/.

Secondly, *WARP should help in choosing the perceived category. With the three constraint families PERCEIVE, *CATEG, and *WARP, we can describe the classification of any acoustic feature value. Suppose, for instance, that a speaker pronounces a front vowel with an F1 of 480 Hz. If everything else is equal, the listener should be more likely to classify this into the nearest category /e/, which is only 80 Hz away, than into the category /ε/, which is 120 Hz away.⁵ This follows automatically from the grammar if the ranking of *WARP is a function of the absolute (two-sided) distance between its arguments x and y , so that *WARP (F1: 480 Hz, 600 Hz) >> *WARP (F1: 480 Hz, 400 Hz):

(4) *Nearest-category classification of an acoustic input of 480 Hz*


[480 Hz]	*CATEG (470) *CATEG (480) *CATEG (490) etc.	PERCEIVE	*WARP (480, 600)	*WARP (480, 400)	*CATEG (400) *CATEG (600)
 /400 Hz/				*	*
/480 Hz/	*!				
/600 Hz/			*!		*
(nothing)		*!			

But this listener, with her symmetric *WARP ranking, does not always perform optimally when considered from a functional viewpoint. Imagine, for instance, that the category /ε/

⁵ This assumes, rather unrealistically, that the perceptual distance is a function of the difference in Hertz. Also, it abstracts away from speaker normalization, which is another function of the perception grammar that enables the listener to classify vowel heights of children (higher F1) and males (lower F1) as well.

occurs three times as often in the language environment as the category /e/. The speaker will produce /ε/ replications with F1 values spread around 600 Hz, and /e/ replications with F1 around 600 Hz. If the distributions of /e/ and /ε/ realizations have the same spreading, then if the speaker pronounces a realization of [500 Hz], i.e. right in the middle between the two categories, it will be more likely that she had meant to produce an /ε/ than an /e/. The best strategy for the listener, then, is to perceive this as the more common /ε/. Because of the frequency difference, the turning point may come to lie at [470 Hz], i.e., a realization of [470 Hz] has an equal probability of stemming from an intended /e/ as from an intended /ε/. The optimal strategy for the listener in this case is to perceive every utterance below 470 Hz as /e/, and every utterance above 470 Hz as /ε/. This must mean that the ranking *WARP (F1: 480 Hz, 600 Hz) and *WARP (F1: 480 Hz, 400 Hz) must be reversed:

(5) *Maximum-likelihood classification of an acoustic input of 480 Hz*

[480 Hz]	*CATEG (470) *CATEG (480) *CATEG (490) etc.	PERCEIVE	*WARP (480, 400)	*WARP (480, 600)	*CATEG (400) *CATEG (600)
/400 Hz/			*!		*
/480 Hz/	*!				
 /600 Hz/				*	*
(nothing)		*!			

By maximizing in this way the likelihood of the intended category, the listener makes sure that for every acoustic input she will make a correct choice (between /ε/ and /e/) in at least 50% of the cases. Thus, this strategy contributes quite directly to the implementation of the functional principle of minimization of confusion. It would be nice if her learning algorithm could automatically lead her to adopt this maximum-likelihood strategy.

2.6 Sequential abstraction

Covert phonological structure, such as features, syllables, or feet, is laid upon the raw acoustic material by the listener's perception grammar. The features were discussed in §2.3. Abstraction of two or more simultaneously occurring features into larger units is handled by *path constraints* (e.g. *REPLACEPATH; Boersma 1998: §9.11). Abstraction of sequences of perceptual units into larger units is handled by families of *sequential abstraction constraints* (Boersma 1998: chs. 12, 18; Boersma 1999a). The names of these families are abbreviations for Obligatory Contour Principle and Line-Crossing Constraint, the two well-formedness principles from autosegmental phonology:

(6) OCP ($f: x; cue_1 | m | cue_2$):

“A sequence of two acoustic cues cue_1 and cue_2 is perceived as a single value x on the perceptual tier f , **despite** the presence of some intervening material m .”

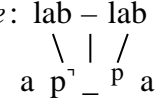

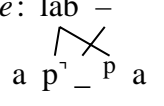
(7) LCC ($f: x; cue_1 | m | cue_2$):

“A sequence of two acoustic cues cue_1 and cue_2 is **not** perceived as a single value x on the perceptual tier f , **because of** the intervening material m .”

2.7 Functionally desirable ranking of sequential abstraction constraints

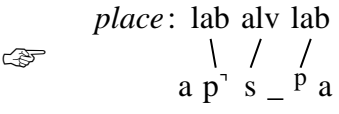
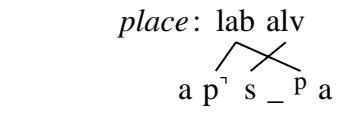
The OCP constraints contribute to a parsimony of perceptual structures, and the LCC constraints contribute to maximizing the maintenance of detailed acoustic information. Compare this with our visual perception of the face of another human being: the OCP tells us to perceive a single face, and the LCC tells us to perceive the nose, the mouth, and the eyes separately. Since a nose usually comes with a mouth and eyes, it has probably been evolutionary advantageous to us to be able to perceive them together as a single higher-level concept: a face. This perceptual integration works in parallel on several levels of abstraction: at the same time, we see nose & mouth & eyes, we see a face, and we see a person. Speech perception works in the same way: if cue_1 and cue_2 go together very frequently, despite some intervening material m , we expect that they will be perceived as one at a relatively low level of abstraction. Evidence for this is found in the phonemic restoration effect (e.g. Samuel 1981, 1990), in which listeners perceive phonemes and syllables continuously across added non-linguistic noises (e.g. coughs); this effect corresponds to the visual integration that takes place if we perceive a person partially obscured by another object. But as with visual integration, phonetic integration will take place with hierarchical organizations as well. For instance, if someone says [apa], we will hear the acoustic sequence $[[a p^{\neg} _ p a]]$, where $[p^{\neg}]$ is the transition (i.e. an unreleased labial stop), $[_]$ is the silence, and $[p]$ is the labial release burst. Quite probably, our language environment is full of intervocalic labial plosives, so the two labial cues will be perceived as a single labial entity on the place tier, despite the intervening silence. We loosely denote the perceptual result as /apa/. This is a case of high-ranked OCP (place):

(8) *Integration of place cues in intervocalic short plosives*

acoustics: $[[a p^{\neg} _ p a]]$	OCP (place: labial; transition silence burst)	LCC (place: labial; transition silence burst)
$place: lab - lab$ 	*!	
 $place: lab -$ 		*

The fact that $[[a p^{\neg} _ p a]]$ will be perceived as /apa/ (placewise) at some level of abstraction does not tell us much if we do not make a comparison with the behaviour of the same acoustic cues in other contexts, i.e. with different intervening material. For instance, the sequence $[[a p^{\neg} s _ p a]]$, with intervening sibilant noise, will probably be perceived with two separate labials on the ‘segment’ level, so that LCC >> OCP this time:

(9) *Integration of place cues in intervocalic short plosives*

acoustics: [[a p ^ɾ _ ^p a]]	LCC (place: labial; transition silence burst)	OCP (place: labial; transition silence burst)
		*
	*!	

Intermediate sequences are more difficult. What if the intervening silence is long? The sequence [[a p^ɾ _^p a]] may be perceived as /ap:a/ (i.e. a single labial value on the place tier) in a language where tautomorphic geminates are common, like Italian, and it may be perceived as /appa/ (two labial values) in a language where a long silence tends to occur on morpheme boundaries, like English. Similar considerations as in this case of geminate plosives can apply in the case of homorganic nasal+plosive clusters. Thus, the ranking of OCP and LCC should depend on the frequency of co-occurrence of the two cues, given the intervening material, and this will roughly mean that the ranking of OCP should be negatively, and the ranking of LCC positively, correlated with the amount of intervening material.

3. The adult recognition grammar

The recognition grammar contains two kinds of constraints: faithfulness and lexical access.

3.1 Faithfulness constraints in recognition

The first group of constraints in the recognition grammar is formed by the faithfulness constraints, which punish any dissimilarities between the perceptual input and the underlying form. The most typical one is:

(10) *REPLACE (*f*: *x*, *y* / *cond* / *left* _ *right*):

“do not recognize an input feature value *x* on the perceptual tier *f* as a different value *y* on the same tier, under the condition *cond* and in the environment between *left* and *right*.”

As an example, consider a language with nasal place assimilation, in which the listener often has to recognize a sequence like /ampa/ as an underlying [an+pa]. This violates a constraint like *REPLACE (place: labial, coronal / nasal / _ consonant), which works on the perceptual place tier, which is categorized into at least two classes (labial and coronal) for nasals.

If the recognition grammar only contained *REPLACE constraints, they could be vacuously satisfied by recognizing an empty underlying form. This is prevented by two correspondence constraints:

(11) RECEIVE (*f: x / cond / left _ right*):

“if the perceptual input contains a feature value *x* on the tier *f*, the underlying form should contain any corresponding value on the same tier.”

(12) DONTRECEIVE (*f: x / cond / left _ right*):

“if the underlying form contains a feature value *x* on the tier *f*, the perceptual input should contain any corresponding value on the same tier.”

These correspondence constraints militate against ignoring phonological material that is available in the input, and against constructing an underlying form for which no phonological material is available in the input. For binary features, *REPLACE can be combined with RECEIVE and DONTRECEIVE (Boersma 1998: §9.8):

(13) *DELETE (*f: x / cond / left _ right*):

“if the perceptual input contains a feature value *x* on the tier *f*, the underlying form should contain the same value on the same tier.”

(14) *INSERT (*f: x / cond / left _ right*):

“if the underlying form contains a feature value *x* on the tier *f*, the perceptual input should contain the same value on the same tier.”

The origin of all these constraints is discussed in §12.1.

3.2 Functionally desirable ranking of faithfulness in recognition

The functional principle by which faithfulness is ranked, is *minimization of perceptual confusion*. For the recognition grammar, this leads to two desirable types of fixed rankings, a language-independent one and a language-dependent one.

The language-independent fixed ranking is based on the detectability of perceptual features in the acoustic signal. The various values of the perceptual place feature, for instance, are detected more easily in plosive consonants than in nasal consonants: /p/ and /t/ are better distinguishable than /m/ and /n/, as is generally known by many who have tried to spell their names over a telephone line, and is confirmed by confusion matrices from listening experiments that measure identification of phonemes in noisy conditions (e.g. Pols 1983). Since the place cues of /p/ and /t/ are more reliable than those of /m/ and /n/, it will be advantageous for the listener if she takes a perceived /p/ and /t/ as reliably pointing at lexical entries containing |p| and |t|, and if she takes a perceived /m/ and /n/ with a grain of salt. This means that her constraint ranking must be such that a perceived /p/ is preferably mapped onto an underlying |p|, but that a perceived /m/ may have a chance of being mapped onto an underlying /n/. In constraint language:

(15) *Fixed ranking based on acoustic similarity*

*REPLACE (place: lab, cor / plosive) >> *REPLACE (place: lab, cor / nasal)

The other, language-dependent, fixed ranking is based on the frequency of occurrence of the various feature values (Boersma 1998: §9.5). In most languages, coronal consonants occur

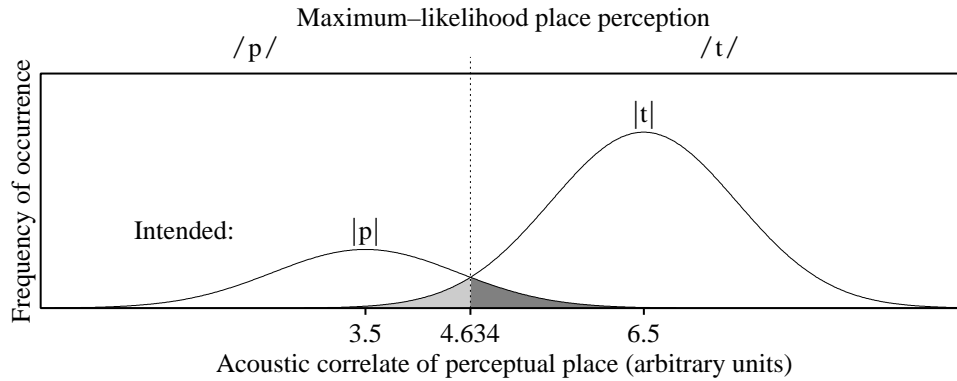


Fig. 4 Two consequences of a frequency difference between labials and coronals: a category shift in perception, and a correction bias in recognition

much more frequently in utterances than labial consonants. On the perceptual place tier in the *perception grammar*, this would lead to a shift of the labial/coronal category boundary in the labial direction (Figure 4, adapted from Boersma 1998: 181), much like the vowel height shift discussed above in §2.5. What concerns us here is the influence on the *recognition grammar*. The maximum-likelihood listener of Figure 4, confronted with a perceived /p/, will have a 9.6 percent probability of having to correct this to an underlying |t|, whereas the probability that a perceived /t/ must be recognized as |p| is only 4.2 percent.⁶ The cognitive representation of this difference in correction probability is

(16) *Fixed ranking based on category markedness*

*REPLACE (place: cor, lab / plosive) >> *REPLACE (place: lab, cor / plosive)

and an analogous fixed ranking will exist for nasals. It would be nice if a learning algorithm could take care of these rankings automatically.

3.3 Lexical-access constraints

The second group of constraints in recognition are constraints against lexical access:

(17) *LEX (*x* / *context*):

“do not recognize an utterance as the lexical item *x* in the given semantic *context*.”

For instance, *LEX (|hɛə| ‘hair’ / ‘cut’) is a constraint against recognizing anything as the lexical item |hɛə| ‘hair’ in the semantic context of cutting.

The origin of these constraints is discussed in §9.

3.4 Functionally desirable ranking of *LEX

To minimize the probability of misrecognition, the listener should rank her *LEX constraints higher for less frequently occurring words or for semantic contexts that are less applicable in the current communicative situation.

⁶ Computation with Bayes: $P(|t| \text{ given } /p/) = P(/p/ \text{ given } |t|) \cdot P(|t|) / P(/p/) = (\text{light shaded area}) / (|t| \text{ area}) \cdot 3/4 / 0.241 = 0.096$.

If a perceived /hɛə/ means |hɛə| ‘hair’ 90 percent of the time and |hɛə| ‘hare’ 10 percent of the time, it is advantageous for the listener (*multis ceteris paribus*) to recognize it as |hɛə| ‘hair’ all of the time. Thus, if the semantic context is neutral, *LEX (|hɛə| ‘hare’) should outrank *LEX (|hɛə| ‘hair’).

In real life, the semantic context tends not to be neutral. It is advantageous for the listener, therefore, to take it into account, so that /hɛə/ will be recognized as |hɛə| ‘hare’ in the sentence ‘we saw /hɛəz/ jumping about in the fields’. Thus, *LEX (|hɛə| ‘hair’ / ‘fields’) should outrank *LEX (|hɛə| ‘hare’ / ‘fields’).

We will see in §11.2 and §11.3 that the Gradual Learning Algorithm will automatically take care of these rankings, even though the context-dependent ranking may also directly reflect extralinguistic knowledge representations.

3.5 Interaction between faithfulness and lexical access

As an example, taken from Boersma (1999b), consider the case in which a Dutch speaker pronounces the lexical item |ɾad| ‘wheel’ as [ɾat], quite appropriately with devoicing of the final obstruent, in an appropriate semantic context that we can denote by the verb ‘turn’. The listener will perceive this as /ɾat/, and her task, now, is to recognize this as |ɾad| ‘wheel’. There are two obstacles for her correct recognition. First, Dutch also has a word |ɾat| ‘rat’, which is phonologically closer to the perceived form. Thus, recognizing /ɾat/ as |ɾad| ‘wheel’ violates the faithfulness constraint *REPLACE (voice: -, +), or simply *INSERT (+voice). It is the semantic context ‘turn’ that must override this phonological preference. Secondly, Dutch also has a common word |vil| ‘wheel’ next to |ɾad| ‘wheel’. Thus, the listener may expect the word |vil| ‘wheel’ more than |ɾad| ‘wheel’ in the given semantic context ‘turn’. It is the small phonological difference between /ɾat/ and |ɾad| that must override this semantic preference. The following tableau gives a formalization of all these interactions:

(18) *Recognizing the wheel*

/ɾat/ context = ‘turn’	*LEX (ɾat ‘rat’ / ‘turn’)	*REPLACE (height)	*LEX (ɾad ‘wheel’ / ‘turn’)	*LEX (vil ‘wheel’ / ‘turn’)	*INSERT (+voice)
ɾat ‘rat’	*!				
☞ ɾad ‘wheel’			*		*
vil ‘wheel’		*!		*	*

This tableau is to be understood as follows. The first candidate loses because the semantic context ‘turn’ disfavors the recognition of rodent animals, and favors the recognition of things likely to turn. This semantic preference is given by the ranking

$$*LEX (|ɾat| ‘rat’ / ‘turn’) \gg *LEX (|ɾad| ‘wheel’ / ‘turn’)$$

This semantic preference outranks the phonological preference for the faithful form |ɾat| ‘rat’, which is given by the violation of *INSERT (+voice). This constraint must be ranked low, because underlyingly voiced obstruents are routinely devoiced in Dutch, so that

listeners must be able to routinely add a [+voice] specification on final obstruents during recognition. For the third candidate, the semantic preference, based on frequency of occurrence, is given by the ranking

*LEX (|ɾɑd| ‘wheel’ / ‘turn’) >> *LEX (|vɪl| ‘wheel’ / ‘turn’)

In this case, the semantic preference is outranked by the phonological preference for the less unfaithful form |ɾɑd| ‘wheel’, because a recognition of /ɾɑt/ as |vɪl| ‘wheel’ would violate a number of high-ranked faithfulness constraints, among which *REPLACE (height: low, high). This constraint must be ranked high because Dutch speakers have high-ranked vowel-height faithfulness, so that in order to minimize confusion listeners should rely on the surface vowel height during recognition.

4. The adult production grammar

Like the listener’s recognition grammar, the speaker’s production grammar contains faithfulness constraints that are in interaction with constraints that evaluate the output candidates.

4.1 Faithfulness in the production grammar

Faithfulness constraints in the production grammar punish dissimilarities between the underlying form and the perceptual output. These constraints are the exact mirror images of the faithfulness constraints in the recognition grammar:

(19) *REPLACE (*f*: *x*, *y* / *cond* / *left _ right*):

“do not realize a feature value *x* on the perceptual tier *f* in the underlying form as something that you will perceive as a different value *y* on the same tier in the perceptual output.”

If there were only *REPLACE constraints, they could be vacuously satisfied by not producing any sound at all. This is prevented by two correspondence constraints (Boersma 1998: 176):

(20) TRANSMIT (*f*: *x* / *cond* / *left _ right*):

“if the underlying form contains a feature value *x* on the tier *f*, the perceptual output should contain any corresponding value on the same tier.”

(21) DONTTRANSMIT (*f*: *x* / *cond* / *left _ right*):

“if the perceptual output contains a feature value *x* on the tier *f*, the underlying form should contain any corresponding value on the same tier.”

These correspondence constraints militate against suppressing phonological material that is available underlyingly, and against producing a perceptual output for which no phonological material is available underlyingly. For binary features, *REPLACE can be combined with TRANSMIT and DONTTRANSMIT (Boersma 1998: §9.8):

(22) *DELETE (*f: x / cond / left _ right*):

“if the underlying form contains a feature value *x* on the perceptual tier *f*, the perceptual surface form should contain the same value on the same tier.”

(23) *INSERT (*f: x / cond / left _ right*):

“if the perceptual surface form contains a feature value *x* on the tier *f*, the underlying form should contain the same value on the same tier.”

The origin of all these constraints is discussed in §13.1.

4.2 Functionally desirable ranking of faithfulness in production

The ranking of the faithfulness constraints in the production grammar should usually reflect the ranking in the recognition grammar. After all, speakers will easily get away with suppressing features that listeners can easily reconstruct. The correspondence between the faithfulness constraints in the two grammars is as follows:

(24) *Correspondence of faithfulness constraints in recognition and production*

Recognition grammar	Production grammar
*REPLACE (<i>f: x, y</i>)	*REPLACE (<i>f: y, x</i>)
RECEIVE (<i>f: x</i>)	DONTTRANSMIT (<i>f: x</i>)
DONTRECEIVE (<i>f: x</i>)	TRANSMIT (<i>f: x</i>)
*DELETE (<i>f: x</i>)	*INSERT (<i>f: x</i>)
*INSERT (<i>f: x</i>)	*DELETE (<i>f: x</i>)

For instance, the fixed rankings (15) and (16) of the recognition grammar should correspond to the following fixed rankings in the production grammar:

(25) *Fixed ranking of faithfulness in the production grammar*

*REPLACE (place: cor, lab / plosive) >> *REPLACE (place: cor, lab / nasal)
 *REPLACE (place: lab, cor / plosive) >> *REPLACE (place: cor, lab / plosive)

Thus, if the listener has little trouble recognizing a perceived /p/ as |t|, the speaker will have few hesitations pronouncing an underlying |t| as /p/.

4.3 Articulatory constraints

As in the recognition grammar, the faithfulness constraints are in interaction with constraints that tend to cause violation of faithfulness. In the case of the production grammar, these are the *articulatory constraints*:

(26) *GESTURE (*a: g / distance, duration, velocity, precision*):


“a certain articulator (or combination of articulators) *a* does not perform the gesture *g*, over a certain *distance*, during a certain *duration*, and with a certain *velocity* and *precision*.”

For instance, an articulatory gesture in [ata] is a closing and opening of the tongue tip, so the articulation [ata] simply violates the constraint *GESTURE (tongue tip: close & open).

4.4 Functionally desirable ranking of articulatory constraints

Humans have evolved in an environment with limited food resources. As a consequence, they have an innate desire to minimize the effort associated with their actions. Therefore, if everything else (i.e. the probability of confusion) is equal, speakers should choose the least effortful implementation of the given perceptual specification. This desirable property of the production grammar is guaranteed if the ranking of articulatory constraints closely mirrors their relative articulatory effort. For instance, if the speaker judges that uvular trills are easier to make than apical trills, she will represent the relevant effort difference as the ranking *GESTURE (tongue tip: up / quite precise) >> *GESTURE (tongue body: back & up / quite precise), or as the somewhat imprecise shorthand *[r] >> *[R],⁷ and if an underlying form contains the perceptual feature |trill|, this speaker will produce a uvular trill:

(27) *Faithful implementation of a trill*

trill	*DELETE (trill)	*[r]	*[R]
[r] /trill/		*!	
 [R] /trill/			*
[s] /sibilant/	*!		
[ʏ] /fricative/	*!		

Note the double representation in each candidate cell, which is a combination of the articulatory and the perceptual output of the production grammar. In the first candidate cell, the notation [r] is a shorthand for the articulatory form [tongue tip: up & precise; velum: raised; glottis: adducted; lungs: contracting; lips: open], whereas /trill/ refers to the perception of low-frequency vibration. The articulatory constraints *[r] and *[R] evaluate the articulatory forms, whereas the faithfulness constraint *DELETE (trill) evaluates the perceptual similarity between the underlying form |trill| and the perceptual forms.


4.5 Interaction between faithfulness and articulatory constraints

Consider the case of nasal place assimilation, in which all nasals assimilate to any following consonant, but plosives do not. This situation, but not the reverse situation in which plosives assimilate but nasals do not, is allowed by the fixed ranking (25a). For nasal place assimilation to work, the two faithfulness constraints must sandwich the articulatory constraint that would be satisfied by assimilation. In the case of an underlying sequence

⁷ It is imprecise because the notations *[r] and *[R] seem to refer to the effort associated with producing the trills, whereas the more precise formulation only refers to the relevant tongue gestures, correctly disregarding all the other gestures needed for the implementation of a voiced trill, such as lung contraction, glottal adduction, velum raising, and lip opening, whose *GESTURE constraints are probably ranked lower.


|an+pa| pronounced as [ampa] (shorthand for bilabial closure, velum raising, etc.), the articulatory gain is the loss of an entire coronal closing and opening gesture (as compared with [anpa]), so the relevant articulatory constraint is *GESTURE (tongue tip: close & open). The speaker will perceive [ampa] as /ampa/ (shorthand for nasality, labial place, etc):

(28) *Nasals assimilate*

an+pa	*REPLACE (cor, lab / plosive)	*GESTURE (tongue tip)	*REPLACE (cor, lab / nasal)
[anpa] /anpa/		*!	
 [ampa] /ampa/			*

Plosives will be pronounced faithfully:

(29) *Plosives do not assimilate*

at+ma	*REPLACE (cor, lab / plosive)	*GESTURE (tongue tip)	*REPLACE (cor, lab / nasal)
 [atma] /atma/		*	
[apma] /apma/	*!		

Many more examples can be found in Boersma (1998).

4.6 The difference between markedness constraints and articulatory constraints

Superficially, our model of interacting articulatory and faithfulness constraints resembles the traditional Optimality-Theoretic distinction between markedness constraints and faithfulness constraints. Functionally speaking, however, markedness can not only be related to articulatory effects, which are expressed through gestural constraints, but also to perceptual effects, which are expressed through faithfulness constraints.

For instance, the fact that /s/ is much more common cross-linguistically than /θ/, is typically accounted for in generative OT by a ranking like *θ >> *s. For example, see Prince & Smolensky (1993: 181: *LAB >> *COR). But if we look at the relative articulatory effort associated with these sounds, we must note that [s] involves a rather complicated tongue-grooving gesture, which is needed to create the jet of air that generates a sibilant noise when it hits the teeth ridge (Hardcastle 1976; Stevens 1998). No such complicated gesture is needed for [θ], so a priori we probably have the ranking *GESTURE (tongue tip: approach & groove) >> *GESTURE (tongue tip: approach & non-groove), which we could abbreviate as

(30) *Articulatory markedness*

*[s] >> *[θ]

I have made this ranking stand out in white, because it is so very different from traditional OT. On the basis of this ranking alone, we expect that /θ/ should be more common than /s/ cross-linguistically. The question, of course, is why the reverse is true. The answer is that the

noise of /s/ is 10 to 15 dB louder than that of /θ/, which makes /s/ much better audible than /θ/ in a world with background noise. So on the basis of perceptual salience we probably have another a priori ranking:

(31) *Perceptual markedness*

TRANSMIT (/s/) >> TRANSMIT (/θ/)

The most common system of front coronal fricatives, namely the system consisting of /s/ only, results from the ranking

(32) *Why /s/ is more common than /θ/*

TRANSMIT (/s/) >> *[s] >> *[θ] >> TRANSMIT (/θ/)

The reader may note a problem with factorial typology. It would seem that (30) and (31) allow the ranking

(33) *A markedness reversal*

*[s] >> TRANSMIT (/s/) >> TRANSMIT (/θ/) >> *[θ]

which would be a language with /θ/ but without /s/. The cause of the commonness of (32) and the utter rarity of (33) is the fact that adult speakers, with fully developed motor skills and after years of training, will find [s] not much more effortful than [θ], whereas the difficulty of hearing /θ/ in background noise will never go away, so that the advantage of /s/ over /θ/ will stay until old age causes the loss of high-frequency hearing. The ranking of (33) is expected only in cases where the speaker has difficulties with complicated articulations. This occurs mainly during the early years, when many children pronounce an intended [s] as [θ]. Speech therapists have earned livings trying to lower *[s].

4.7 Positional faithfulness versus positional markedness

There are often two ways to describe positional neutralization in OT. Nasal place assimilation, for instance, has been described as the ranking NASASSIM >> IDENT-IO(place), where NASASSIM means “a nasal consonant should have the same place as a following consonant”, and IDENT-IO(place) expresses place faithfulness (e.g. Padgett 1995). Since NASASSIM can be rephrased as “no branching place within a nasal-plus-consonant cluster”, it is an instance of the *STRUC family (Prince & Smolensky 1993), also called a structural or markedness constraint. If we add a constraint against branching place within plosive-plus-consonant clusters, place assimilation confined to nasals can be expressed as NASASSIM >> FAITH >> PLOSASSIM, i.e. a positional markedness sandwich. By contrast, (28) expresses nasal place assimilation as a positional faithfulness sandwich.

In generative OT, people have tried to argue that positional neutralization is due to position-dependent ranking of faithfulness (e.g. Jun 1995, Beckman 1998) or to position-dependent ranking of markedness (e.g. Zoll 1998). From a functional viewpoint, however, markedness, in the sense of cross-linguistic rarity, emerges from an interaction between principles of articulatory effort and perceptual confusion, and is not encoded directly in our phonological language device (§4.6). Still, functional phonology has a comparable distinction

between faithfulness constraints and articulatory constraints. The question, then, is whether positional neutralization is due to positional ranking of faithfulness or to positional ranking of articulatory constraints. From the viewpoint of the functional principles involved, the answer depends on whether it is perceptual confusion or articulatory effort that is position-dependent in each case. For nasal place assimilation, the articulatory gain involved in assimilating away a tongue-tip gesture is equal for nasals and for plosives, so this must be a clear case of positional faithfulness. Examples of positional articulatory effort will be harder to find (for adult speakers), because articulatory gestures tend to lose their distinction in terms of effort, whereas the probability of perceptual confusion will always stay dependent on loudness and acoustical similarity (§4.6).

5. The initial state of the grammar

With a knowledge of the constraints in the three grammars and their functionally desirable rankings, we can now turn to the subject of the acquisition of these constraints and their rankings.

According to most Optimality-Theoretic literature on phonological acquisition (Gnanadesikan 1995, Smolensky 1996a), the child's grammar starts out with a set of innate structural constraints (which disfavour marked phonological structures) dominating a set of innate faithfulness constraints (which favour similarity of the output to the input). The acquisition process would then only have to rerank all these innate constraints with respect to one another.

However, many of the innate structural constraints that have been proposed refer explicitly to specific phonological features and feature values, which must then also be innate. For instance, the constraint NASASSIM, which says that nasal consonants must have the same place of articulation as the following consonant, explicitly refers to the features [place] and [nasal], and to the feature value [+nasal]. The problem with these innate structural constraints, then, is that they presuppose innate phonological representations, for which there seems to be no evidence (§8.2).

The alternative proposal, defended here, is that the production grammar initially contains no constraints at all. Above, I showed that the actions of NASASSIM can be handled quite well by the structural constraint *GESTURE (tongue tip: close & open), in interaction with the fixed ranking of the faithfulness constraint *REPLACE (place / plosive) above the faithfulness constraint *REPLACE (place / nasal). These constraints are members of very general families, and are created by the learner herself. Thus, the articulatory constraint *GESTURE (tongue tip: close & open) enters the child's production grammar (initially high-ranked) as soon as the child learns (by play, i.e. practising sensori-motor relations) that this tongue gesture implements the perceptual goal of producing a coronal stop. Likewise, the faithfulness constraint *REPLACE (place / nasal) enters the grammar (initially low-ranked) as soon as the child has acquired the perceptual tiers of nasality and place and divided these tiers into some categories. Finally, the fixedness of the ranking of *REPLACE (place / plosive) above *REPLACE (place / nasal) is a result of the general dependence of the ranking of faithfulness on the amount of perceptual confusion (the place of plosives is easier to determine than the place of nasals).

6. Learning algorithm

As we will see below, many phenomena in phonology are explained as results of the actions of a deaf, dumb, and blind learning algorithm. According to this gradual constraint-ranking learning algorithm (Boersma 1997, 1998: ch. 14, to appear; Boersma & Levelt 1999; Boersma & Hayes 1999), the child will take action as soon as she discovers that an adult form is different from produced by her own grammar (this discovery is a result of the ‘comparison’ in Figure 1). Her action will consist of raising all the constraints that are violated by her incorrect winning candidate, by a small step up the continuous ranking scale. If the adult form happens to be in her list of candidates as well (i.e., if she knows about one or more articulations that will lead to a perceptual result equal to the adult form), she will also lower all the constraints that are violated in her grammar by the correct candidate, by the same small step down the ranking scale.

We will see many examples in the following sections.

7. Formal details of acquisition

The child will try to match as well as possible what she hears in her environment. Differences between her speech and the adult’s come about as results of the non-final states in her acquisition of (1) perception, (2) the lexicon, and (3) production.

At every point during the acquisition of language, the child entertains her own production, perception, and recognition grammars, gradually heading toward a language that resembles the one in her environment. Changes in these grammars are brought about by three innate devices: the mammalian Perception Acquisition Device (PAD), the human Language Acquisition Device (LAD), and the vertebrate Gradual Learning Algorithm (GLA). In the following, I will give a simplified sketch of how these devices control the acquisition of phonology.

Initially, the perception and production grammars are empty, and the recognition grammar contains a single constraint *LEXICALIZE, which militates against the storage of a new lexical item. Then, the following seven processes will occur in parallel:

- (34) *The seven parallel processes of the acquisition of phonology*
- a. As soon as a feature value is perceived, PAD will supply the perception grammar with constraints against its categorization and with faithfulness constraints. GLA will automatically lead to a realistic probability-matching maximum-likelihood criterion (§8).
 - b. As soon as a perceptual category is created, PAD will supply the perception grammar with constraints that control perceptual abstraction (OCP and LCC). When entering the perception grammar, OCP constraints are high ranked, whereas LCC constraints are low ranked. GLA will automatically rank these according to frequency: two acoustic cues or perceptual events that occur in sequence more frequently will develop a larger probability of being analysed as a single more abstract percept (§9).

- c. The recognition grammar forces the creation of a new lexical item as soon as each candidate that is already in the lexicon, is either phonologically or semantically too dissimilar from the perceived adult form. Technically speaking, the winning candidate is a new lexical item, equal to the adult form; it violates *LEXICALIZE, and wins because every other candidate violates a higher-ranked *LEX or FAITH constraint (§10).
- d. As soon as a lexical item is created, LAD supplies the recognition grammar with a constraint (*LEX) against its recognition. The semantic contexts contribute dynamically and additively to the ranking of this constraint. GLA will automatically rank this lexical access constraint according to frequency: the recognition of more frequently occurring words will be preferred. Furthermore, GLA will automatically determine the weight with which each semantic context contributes to the ranking of each *LEX constraint (§11).
- e. As soon as a perceptual category is created, LAD supplies the recognition grammar with several faithfulness constraints (FAITH), which favour the similarity between the perceived form and the lexical form (with respect to the presence of features, their co-occurrence, and their sequencing). GLA will automatically rank these constraints according to the frequency of occurrence of the associated phonological feature values (§12).
- f. LAD forces the use of the *same* faithfulness constraints, perhaps with the same ranking, in the production grammar as well as in the recognition grammar (§13).
- g. As soon as the child learns the perceptual result of an articulatory gesture, LAD creates an articulatory constraint (*GESTURE) at the top of the production grammar. GLA will usually lower this constraint, and automatically raise the corresponding faithfulness constraints, thus reranking the constraints in the direction of a grammar that produces a more adult-like output (§14).

The conclusion could be that the innate LAD restricts itself to introducing constraints of a very general nature (faithfulness, lexical access, gesture), and to forcing the use of a single set of faithfulness constraints shared between the production and recognition grammars.

8. Categorization and acoustics-to-perception faithfulness

The innateness of phonological representations is an assumption without which much of current phonological theory could not exist. For instance, theories of feature geometry presuppose the existence of a universal-and-arbitrary (therefore, innate) feature geometry, and these theories see it as their task to find this geometry. As another example, OT feature phonology has proposed substantive innate constraints like NASASSIM, which must presuppose the existence of an innate feature [nasal]. However, I will show that there is no support for innate feature values outside these theories, and that feature values can be learned by a general innate phonetic categorization system.

8.1 Many categories are learned

Children are born with the capability of distinguishing many more phonetic differences than they will need when speaking their future language. Between the ages of 6 and 12 months, this capability is largely replaced by a *categorical perception* of the distinctions used in their specific language environment, i.e., older infants will start to perceive two stimuli from the same category as *the same* and two stimuli from opposing sides of the category boundary as *different*. In an investigation on the perception of the consonant place continuum, Werker & Tees (1984) found that younger English-learning infants could distinguish the English [b̥a] and [ɖa], the Hindi [t̪a] and [ta], and the Nthlakapmx [kʰi] and [qʰi], whereas older English-learning infants could only distinguish the English pair. In an investigation on the perception of the vowel place continuum, Polka & Werker (1994) found that younger English-learning infants could distinguish German [ʏ] from [u] and German [y] from [u], and older ones could not.

The evolutionary advantage of categorical perception in the communicative situation is clear: in a world where variation is the norm, the listener is required to be able to map many acoustically different events on the same discrete units in the lexicon, so it is often advantageous not to distinguish between sounds that are acoustically close. Thus, English-learning infants may merge the Hindi place contrast because dental stops in English are positional variants of alveolar stops and are best perceived as alveolars, and they merge the Nthlakapmx contrast because uvular stops in English should best be considered occasional realizations of velars. Likewise, both German [y] and [u] are within the range of English /u/, which is often fronted or unrounded (both fronting and unrounding have a raising effect on the second formant). For non-native sounds that do not map to anything in the child's language, the original acoustically based perception seems to stay intact, though the evidence is weak. Thus, English-learning children (and English-speaking adults) can still distinguish the Zulu lateral and medial click (Best, McRoberts & Sithole 1988), possibly because it is not advantageous for the English listener to be able to map either of these sounds to an English phoneme (although it might also be the case that English speakers perceive them categorically, because these sounds do occur as less-linguistic utterances for these speakers).

These results fit in well with categorization in other areas of cognition. Rather than having innate categories for predators like [tiger], [lion], and perhaps [wolf], humans have developed a general innate instrument for biological taxonomy. Infants have been shown (Quinn, Eimas & Rosenkrantz 1993) to be able to learn the demonstrably non-innate category [dog], and I can testify that Dutch 6-to-8-year-olds have no trouble keeping track of the rapidly evolving subspecies of *pokémon*. In a changing world, generic categorizers have an evolutionary advantage. In a world in which the newborn does not know which language she will encounter, she is at an advantage if she is capable of constructing new categories based on the data she hears.

8.2 Category boundaries present from birth only occur at acoustical discontinuities

We saw that infants are capable of categorical perception, and that some of the categories (for English: coronal place, dorsal place, rounded high vowel) are learned in the second half year. Relevant for the question whether phonological theory should accommodate innate substance (feature geometry, NASASSIM constraints), however, is the question whether there are also

innate categories. If some categories are innate, we expect that infants have them in the first half year without training and that all adults have identical categories or at least identical category boundaries (which may be different from the infants' because of maturation of the auditory system). If all categories are learned, we expect that all of them develop in a language-particular way during the first years, and that languages vary to a large extent with respect to the location of category boundaries, at least for those features that correspond with acoustically continuous scales.

Young infants do seem to show effects of untrained categorical perception. Eimas, Siqueland, Jusczyk & Vigorito (1971) found that very young infants categorized the p–p^h part of the VOT continuum (i.e. the English “b”–“p” continuum) in a way reminiscent of the clustering of VOT values that Lisker & Abramson (1964) found in the languages of the world. This would be a clear case for the innateness of category boundaries. But there are two problems.

First, VOT is not an acoustically continuous scale, since the perceptual difference between /b/ and /p/ is basically that of voicing murmur versus silence, whereas the difference between /p/ and /p^h/ boils down to voiced (i.e. loud) formant movements versus aspiration noise. So we expect that there are perceptual discontinuities along this scale (in the case of Eimas et al., a discontinuous distinction between the stimuli could be whether the first formant was heard simultaneously with the second and third formants or not), and that discrimination may be better across these discontinuities. This was confirmed by Kuhl & Miller (1978), who showed that chinchillas hear a perceptual boundary at the same location as human infants. After these animals were trained to discriminate a 0-ms and a 80-ms VOT, their perceptual boundary was not just somewhere in the middle, but around 25, 34, and 42 ms for /p^h/, /t^h/, and /k^h/, respectively, just as for adult listeners of English (who had been shown by Eimas et al. to behave as infants do). In line with this, Eggermont (1995) investigated neural patterns in the auditory cortex of untrained anaesthetized cats, and found that for stimuli with short VOT, neurons tended to show a single burst of activity (corresponding to the release burst of [p]), and that for stimuli with long VOT, there were two activity bursts (corresponding to the release burst and the voice onset of [p^h]). To be true, the fact that each neuron had its own ‘perceptual boundary’, almost uniformly distributed throughout the entire 10–70 ms region, led Eggermont to conclude that it is unlikely that it is the auditory cortex of the cat that causes categorical perception of VOT. Nevertheless, a connectionist would add a trivial single-neuron integrator, and thus have these cortical neurons decide the single/double issue on the basis of a majority vote, which would produce a robust boundary near 40 ms.

The second problem is that there is no evidence for clustering if we look at more languages than Lisker & Abramson did. Cho & Ladefoged (1997), measuring the p–p^h part of the VOT continuum in 17 languages, found a much more continuous distribution. This means that there are many languages that have the VOT boundaries at different locations from the infants. Thus, many children would have to make two modifications to the original three-way categorization: collapsing two of the three categories into one, and moving a category boundary to a language-specific location. The former should make us doubt the need for innate cognitive feature values, the latter should make us doubt the need for innate phonetic category boundaries.

A close look at exactly what kinds of perceptual dimensions show early language-independent categorization (place, voice onset time) reveals that these dimensions create discontinuities already at the peripheral auditory level. By contrast, perceptual dimensions that are continuous at the peripheral auditory level (tone, vowel height) are not perceived categorically by young infants (Swoboda, Morse & Leavitt 1976), and show language-specific class boundaries as soon as any categorization emerges.

All research on the development of perception in infants (for reviews, see Vihman 1996; MacNeilage 1997; Jusczyk, Houston & Goodman 1998) seems therefore compatible with the view that early untrained categorization, if present, corresponds to auditory capacities general to mammals (and, therefore, not optimized for speech), and that language-specific phonological features and structures are learned from the environment by an innate generic phonological categorizer, and not given as innate cognitive or perceptual categories.

8.3 Origin

(34a) *Origin of categorization and acoustics-to-perception faithfulness constraints*

As soon as a feature value is perceived, the Perception Acquisition Device will supply the perception grammar with constraints against its categorization and with faithfulness constraints.

For this to work, we must assume that the perceptual tier (itself is already in place. Whether or not any tiers are innately given, is a question for which there is much less evidence than for the question whether or not the *values* along these tiers are innate, and I will not try to review it here. Thus, when the learner, already equipped with a first-formant tier, hears an F1 of 450 Hz for the first time, she will create a constraint PERCEIVE (F1: 450 Hz), a constraint *CATEG (F1: 450 Hz), and a family of constraints *WARP (F1: 450 Hz, x).

8.4 Ranking


(34a) *Ranking of categorization and acoustics-to-perception faithfulness constraints*

GLA will automatically lead to a realistic probability-matching maximum-likelihood criterion.

The second and third desirable properties of categorization discussed in §2.5 (classification into the nearest category, and a maximum-likelihood criterion) are automatic results of the learning algorithm.

Suppose that the intended /400 Hz/ and /600 Hz/ categories are equally common in the speaker's utterances, but that the *WARP (F1) constraint family is ranked completely incorrectly, e.g. *WARP (F1: 420 Hz, 400 Hz) >> *WARP (F1: 420 Hz, 600 Hz), so that any acoustic input of [420 Hz] is classified into the /600 Hz/ category instead of into the much nearer /400 Hz/ category. In the usual case that the speaker's intention was to transmit the /400 Hz/ category, this leads to a misperception:

(35) *Wrong classification of an acoustic input of 420 Hz*


[420 Hz] intended: /400 Hz/	*CATEG (other)	*RECEIVE	*WARP (420, 400)	*WARP (420, 600)	*CATEG (400)	*CATEG (600)
√ /400 Hz/			*!→		*→	
/420 Hz/	*!					
*  */600 Hz/				←*		←*
(nothing)		*!				

For learning to be possible, we must assume that the learner notices the mismatch, i.e., she comes to know that although she perceived /ε/, the speaker’s intention was /e/. We represent this awareness in tableau (35) by mentioning the apparently correct category /400 Hz/ in the top left cell, i.e. as data that the learner has about the adult utterance. The learner will mark her own winner (the winning candidate in her own grammar) as incorrect, which is depicted with the asterisks around the pointing finger in the tableau. The form that the learner regards as correct also appears in the tableau, and receives a check mark (√).

As a result of the mismatch, the learner will take a learning step, which consists of demoting the constraints that are violated in the adult form (denoted by the rightward arrows) and promoting the constraints that are violated in her own form (as shown by the leftward arrows). Gradually, after many of these mismatches, the two *WARP constraints will approach each other, until their ranking is reversed and the learner will perceive the /400 Hz/ category. After this reversal, the next form that the learner perceives will be equal to the form intended by the speaker; the learner will not notice any discrepancy and see no reason to change her grammar any further.


Beside the gross nearest-category effect, the second effect of the GLA is the establishment of a maximum-likelihood criterion. Suppose, as in §2.5, that the intended category /600 Hz/ is much more common than the /400 Hz/ category, so that the skewed value of [480 Hz] is equally likely to stem from the /600 Hz/ category as from the /400 Hz/ category. Suppose further that the learner’s criterion for the /400 Hz/ – /600 Hz/ decision is right in the middle at [500 Hz], i.e., *WARP (F1: 500 Hz, 400 Hz) is ranked at the same height as *WARP (F1: 500 Hz, 600 Hz). With the fixed rankings within the *WARP family, this means that an acoustic [480 Hz] will always be perceived as /400 Hz/, which is incorrect in 50 percent of the cases:

(36) *50 percent wrong classification of an acoustic input of 480 Hz*

[480 Hz] intended: /600 Hz/	*WARP (480, 600)	*WARP (500, 400) *WARP (500, 600)	*WARP (480, 400)	*CATEG (600)	*CATEG (400)
*  */400 Hz/			←*		←*
√ /600 Hz/	*!→			*→	

In 50 percent of the cases, the intended category was /400 Hz/ and nothing happens. In the other 50 percent of the cases, the two *WARP (480, *cat*) constraints will approach one another, as shown by the arrows in tableau (36). With repeated applications of [480 Hz] utterances, this will continue until the two *WARP constraint have the opposite ranking. After that, a [480 Hz] input from an intended /600 Hz/ will be correctly perceived as /600 Hz/, but a [480 Hz] input from an intended /400 Hz/ (which makes up 50 percent of all [480 Hz] inputs) will be misperceived:

(37) *Again 50 percent wrong classification of an acoustic input of 480 Hz*

[480 Hz] intended: /400 Hz/	*WARP (500, 400)	*WARP (480, 400)	*WARP (480, 600)	*WARP (500, 600)	*CATEG (600)	*CATEG (400)
√ /400 Hz/		*!→				*→
*  */600 Hz/			←*		←*	

This will lead to a new reversal of the two constraints. After this, there will be continual reversals, causing the two constraints to stay in each other's vicinity. The behaviour of this learner surfaces as a tit-for-tat strategy: after every intended /600 Hz/ pronounced as [480 Hz], the learner will perceive the next [480 Hz] as /600 Hz/.

This is where *noisy evaluation* comes in. The point of having a continuous ranking scale (§6) is the permission of having variation in the output: at evaluation time, a small amount of Gaussian noise is temporarily added to the ranking of each constraint, so that two constraints that are ranked at approximately the same height may be sorted in either order during an evaluation (Boersma 1997, 1998: 283, to appear; Boersma & Hayes 1999). Thus, if *WARP (480, 400) is ranked at a height of 31.0 along the continuous ranking scale, and its competitor *WARP (480, 600) is ranked at a height of 32.0, and the standard deviation of the evaluation noise is 2.0, then *WARP (480, 400) still has a chance of 36 percent of outranking *WARP (480, 600) during a certain evaluation (Boersma 1998: 332). This means that if the *plasticity* (the amount by which the rankings change during a learning step) is much smaller than the standard deviation of the noise, the learner will not show tit-for-tat behaviour. For instance, if the plasticity is 0.1 (i.e. only five percent of the standard deviation of the noise), the ranking difference between the two *WARP constraints will hover between -0.4 and +0.4, so that the probability that [480 Hz] is perceived as /400 Hz/ will hover between 44% and 56%, with a moderate recency effect.

The noisy evaluation is also necessary to explain the learner's behaviour far away from the turning point of [480 Hz]. Consider the [540 Hz] utterances. Suppose that these have a 90 percent probability of being the result of an intended /600 Hz/ category, and 10 percent of stemming from a /400 Hz/ intention. If the learner perceives all of these utterances as /600 Hz/, she will be correct in 90% of the cases, but she will be incorrect in 10%, and in these rare cases she will adapt her grammar by raising *WARP (540, 600) and lowering *WARP (540, 400). If there were no evaluation noise, this reranking would proceed until the relative ranking of the two constraints is reversed. This would upset the idea that *WARP is ranked according to the distance between the perceptual and the acoustic form. But the evaluation noise comes to the rescue: if *WARP (540, 600) rises to, say, a distance of 3.0

below *WARP (540, 400), there will be a probability of 14% that *WARP (540, 600) outranks *WARP (540, 400) at the next evaluation. This means that two probabilities will compete: first, the chance that *WARP (540, 600) must be *lowered* at the next evaluation equals the chance that the intention was /600 Hz/ (i.e. 90%) times the chance that the perception will be /400 Hz/ (i.e. 14%), which is 12.6%; second, the chance that *WARP (540, 600) must be *raised* at the next evaluation equals the chance that the intention was /400 Hz/ (i.e. 10%) times the chance that the perception will be /600 Hz/ (i.e. 86%), which is 8.6%. Thus, if the plasticity is 0.1, the expected change in the ranking of *WARP (540, 600) is a lowering of $(12.6\% - 8.6\%) \cdot 0.1 = 0.004$. So we see that if *WARP (540, 600) happens to rise to a distance of 3.0 below *WARP (540, 400), it is expected to fall again. An equilibrium will arise when the distance is 3.6: the learner will then have a 90% chance of perceiving an input [540 Hz] as /600 Hz/, and a 10% chance of perceiving it as /400 Hz/. In such a case, the chance of rising at the next evaluation ($10\% \cdot 90\% = 9\%$) equals the chance of falling ($90\% \cdot 10\% = 9\%$).

We see that GLA with noisy evaluation leads to an equilibrium in which the listener mimics the frequencies of the input: if the speaker's [540 Hz] utterances have a 10% chance of coming from a /400 Hz/ intention, then the adult listener will have learned to perceive [540 Hz] utterances as /400 Hz/ in 10% of the cases. This *probability matching* behaviour is a practical approximation to the optimal maximum-likelihood listener: the separating criterion (50% cross-over point) between the two categories is equal to that of a maximum-likelihood listener.

8.5 Remaining questions

We didn't discuss where the learner's knowledge of the intended category comes from. In (36) and (37), we see that as a result of the learning process, several *CATEG constraints are reranked as well: the most frequently used categories will get low *CATEG constraints, thus facilitating their perception. It stays unclear, however, how the actual division of F1 into discrete classes, needed for an assessment of the intended category, was brought about. This division can probably not be simulated as an Optimality-Theoretic learning scheme; instead, we could simulate it with any suitable neural-net classification procedure (Grossberg 1976; Carpenter & Grossberg eds. 1991; Behnke 1998) that is told (by the LAD) to handle acoustic and semantic similarity.

Another question concerns the separation of cause and result in the ranking within the *WARP family. There are at least two possibilities. First, we see that GLA causes *WARP (540, 600) to be ranked lower than *WARP (530, 600) because the probability for the speaker to pronounce an intended /600 Hz/ as [540 Hz] is greater than the probability that she pronounces it as [530 Hz]. This would bring the *first* desirable property mentioned in §2.5 under the rule of the GLA as well. But alternatively, we may reformulate *WARP as “do not perceive an acoustic input x as a different value **at least as far away as** y .” In that case, the learning step in (36) will involve a lowering of *WARP (500, 600), which preserves the functional ranking, uninfluenced by the GLA. Note, though, that the preservation is not symmetric, since *WARP (500, 400) does not rise, which becomes problematic once *WARP (480, 400) rises past it.

9. Perceptual abstraction

9.1 Origin

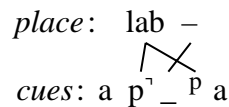
(34b) *Origin of perceptual abstraction constraints*

As soon as a perceptual category is created, the Perception Acquisition Device will supply the perception grammar with constraints that control perceptual abstraction (OCP and LCC).

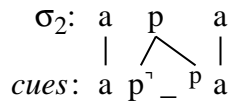
For instance, once the learner has the low-level perceptual categories [a] (high F1), [p̚] (labial implosion), [∅] (silence), and [P] (labial explosion), and she hears the acoustic sequence [[a p̚ _ P a]] for the first time, she will create OCP and LCC constraints for all pairs of acoustic cues, possibly with much intervening material, among which:

(38) *Four integrations from [[a p̚ _ P a]]*

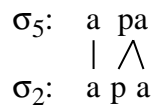
- a. OCP (place: lab | _ | P). If ranked higher than its LCC counterpart, this causes the perception of a single labial value on the place tier:



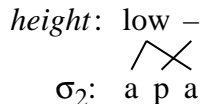
- b. OCP (σ_2 ; p̚ | _ | P). This may cause the perception of a single “segment” /p/, so that [[a p̚ _ P a]] may be /(a)(p)(a)/ on the language-specific second level of abstraction (Boersma 1999a).



- c. OCP (σ_5 : P | | a). This may cause the perception of a single ‘syllable’ /pa/, which may be a unit in the language-specific fifth level of abstraction (Boersma 1999a), so that [[a p̚ _ P a]] may be /(a)(pa)/ on that level:



- d. OCP (height: low; a | p̚ _ P | a). This may cause the perception of /a/ as a single vowel across an intervening consonant, i.e. vowel harmony (Boersma 2000a):



Note the line crossings (LCC violations) in (38a) and (38d). Like the perceptual categories, the abstractions in (38) are unobservable in the articulatory/acoustic surface form. Nevertheless, their psychological reality in many languages is witnessed by the descriptive successes of theories of metrical phonology and autosegmental phonology.



9.2 Initial ranking

(34b) *Initial ranking of perceptual abstraction constraints*

When entering the perception grammar, OCP constraints are high ranked, whereas LCC constraints are low ranked.

If it is true that humans have an innate desire to abstract whenever possible, it is likely that the initial ranking should maximize abstraction, i.e. OCP >> LCC. For instance, this ranking makes sure that /εpε/ is perceived with a single value for vowel height:

(39) *Perception of harmonic vowels*

'segment' level: / ε p ε /	OCP (height; V C V)	LCC (height; V C V)
height: 2 – 2 σ ₂ : ε p ε	*!	
 height: 2 –  σ ₂ : ε p ε		*

Thus, the initial ranking OCP >> LCC tends to simplify hidden structure.



9.3 Language-specific ranking

(34b) *Language-specific ranking of perceptual abstraction constraints*

GLA will automatically rank OCP and LCC according to frequency: two acoustic cues or perceptual events that occur in sequence more frequently will develop a larger probability of being analysed as a single more abstract percept.

If the learner is still in the initial state (39), and she receives a disharmonic form /εpe/, she may have trouble building the correct structure:

(40) *Misperception of disharmonic vowels*

'acoustic' level: [[ε p ¹ _ p e]]	OCP (height; V C V)	*WARP (height: 400, 600)	LCC (height; V C V)
✓ height: 2 – 3 σ ₂ : ε p e	*!→		
height: 2 – 2 σ ₂ : ε p ε	*!	*	
 height: 2 –  σ ₂ : ε p ε		←*	←*

There is simply no candidate that has the correct vowel height and satisfies OCP at the same time. The error in (40) will lead to incorrect lexicalizations and productions. If the learner notices this, she will take action and modify the rankings according to the arrows in (40). Eventually, OCP will fall down until it is below either *WARP (400, 600) or LCC, and the learner will perceive /εpe/ with two different vowels, as appropriate in this language, which apparently shows no tongue-root harmony.

Thus, the initial high ranking of OCP will keep the young learner in a restrictive tongue-root language until she is forced by the data to change this hypothesis. This is the subset principle at work in perception. Should the learner of a tongue-root harmony language inadvertently arrive in a non-harmony hypothesis (LCC >> OCP), she cannot change her hypothesis back. This may seem an unwanted property of our combination of initial state and learning algorithm, but the overt behaviour of this learner will be hard to distinguish from that of a listener who does have harmonic structures (cf. §15.4).

10. Lexicalization


10.1 Origin

(34c) *Origin of lexical items*

The recognition grammar forces the creation of a new lexical item as soon as each candidate that is already in the lexicon, is either phonologically or semantically too dissimilar from the perceived adult form.

Let's return to our Dutch example of §3. Suppose that a child has already stored the lexical items |Rat| 'rat' and |vil| 'wheel', but not yet the less common item |Rad| 'wheel'. Now when she hears an adult produce /Rat/ in the semantic context 'turn', the child may decide that recognizing it as |Rat| 'rat' is inappropriate given the semantic context, and that recognizing it as |vil| 'wheel' is inappropriate given the large phonological unfaithfulness. She will then prefer to create a new item in her lexicon:

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

/Rat/ context 'turn'	*LEX (Rat 'rat' / 'turn')	*REPLACE (height)	*LEXICALIZE	*LEX (vil 'wheel' / 'turn')	*INSERT (+voice)
Rat 'rat'	*!				
vil 'wheel'		*!		*	
 Rat 'wheel'			*		
Rad 'wheel'			*		*

The new item must violate an already available constraint *LEXICALIZE. This constraint cannot be ranked too low; otherwise, the learner would create a new lexical item whenever the semantic context is slightly inappropriate or whenever the adult surface form is slightly different from every existing lexical item.


The winning candidate is |ɾat| ‘wheel’, because every less faithful candidate new item, most notably the adult-like entry |ɾad| ‘wheel’, would violate a faithfulness constraint. The new lexical item will always be equal to the child’s perception of the adult surface form. This automatic result is the functional version of Prince & Smolensky’s (1993: 192) *Lexicon Optimization*. Later on, the learner may reconstruct the new item as |ɾad| ‘wheel’ on the basis of morphological information (perhaps the diminutive /ɾa:ɔ̃ɾɕjə/, cf. /ɾatə/ ‘rat-PL’).

In order to correctly store the new lexical item as |ɾat| ‘wheel’, the learner must know that its meaning is ‘wheel’. Therefore, the learner must know more than the recognizer, who only knows the perceptual input and the semantic context.

10.2 The initial state of the recognition grammar

We can think of the constraint *LEXICALIZE to be available in the recognition grammar from the start, at least before the first lexical item is created. Thus, human beings are born lexicalizers. Suppose that the first lexical item is going to be derived from the adult utterance [ðæʔt], meaning ‘that’. The young child will probably perceive the place of the first consonant, but perhaps not its voicing or frication. Also, she will perceive that the vowel is low. Finally, she may miss the final consonant, since it has much weaker acoustic cues than the first. Suppose, therefore, that the child perceives [ðæʔt] as /ɖæ/. She will store it in her lexicon with the same acoustic image:

(42) *The creation of the first lexical entry*

/ɖæ/	RECOGNIZE	*LEXICALIZE	*REPLACE (height)	*REPLACE (place)
 ɖæ ‘that’		*		
ɖi ‘that’		*	*!	
bæ ‘that’		*		*!
(nothing)	*!			

This tableau is to be understood as follows. The fact that our child accurately perceives consonant place and vowel height means that her grammar already contains faithfulness constraints for these two perceptual features, although they are probably ranked quite low since minimizing confusion by maximizing faithfulness cannot yet be an important drive while the lexicon is still empty.

Finally, the creation of the first lexical item must be forced by a higher-ranked constraint RECOGNIZE, which requires that the perceptual input must not be ignored. This utterly general constraint must be available from the start, as it expresses quite directly the innate drive to communicate by symbols.

10.3 The timing of lexicalization

Before a form can be stored in the lexicon, it must first have been perceived; in this sense, perception predates lexicalization. Before a speaker can use a form in a communicative

situation, she must have stored it in her lexicon; in this sense, lexicalization predates production. But the *development* of perception and lexicalization can go on during the stage in which the child can already speak. In most of the cases, a perceptual contrast has been completely lexicalized once the child starts to implement it in production. For instance, a child acting in an anecdote by Berko & Brown (1960: 531) objected to her mother's pronunciation of [fi] 'fish' as /fɪs/, although she pronounced it as [fis] herself. Likewise, English-learning Amahl (Smith 1973: 3) merged all initial [f] and [w] into [w], but as soon as he could pronounce /f/, he used this new sound in all the words that have [f] in adult English, and in none of the [w] words. However, Amahl did not always show this 'across the board' phenomenon: after he had mastered final [n], he learned to pronounce final [nd], and he used it not only in [wend] 'friend' and [laund] 'round', but also generalized it to [ɸaund] 'brown', which suggests that his lexicon did not yet make a distinction between [n]-final and [nd]-final forms (Smith 1973: pp. 54, 77, 97). For other counterexamples to Smith's claim of Amahl's adult-like lexicalization, see Braine (1976), Macken (1980), and Vihman (1982).

11. Lexical access

Lexical access will be handled by the recognition grammar.

11.1 Origin

(34d) *Origin of lexical access constraints*

As soon as a lexical item is created, the Language Acquisition Device supplies the recognition grammar with a constraint (*LEX) against its recognition. The semantic contexts contribute dynamically and additively to the ranking of this constraint.

When the learner creates the lexical item [ɹat] 'wheel' as a result of tableau (41), she will create a constraint *LEX ([ɹat] 'wheel'). If there are a thousand semantic contexts, including 'turn' and 'bite', there will be a thousand rankings like *LEX ([ɹat] 'wheel' / 'turn') and *LEX ([ɹat] 'wheel' / 'bite'), and nearly half a million rankings with two contexts combined, like *LEX ([ɹat] 'wheel' / 'turn' & 'bite'). Though this may seem like a large number for a single brain to contain (imagine having 10,000 lexical items), the constraint *LEX ([ɹat] 'wheel') can be implemented simply as a single candidate-inhibiting neuron with excitatory or inhibitory connections from 1000 other neurons that represent semantic contexts. In such a view, the ranking of *LEX ([ɹat] 'wheel' / 'turn' & 'bite') is dependent on the rankings of *LEX ([ɹat] 'wheel' / 'turn') and *LEX ([ɹat] 'wheel' / 'bite').

11.2 Ranking by frequency


(34d) *Ranking of lexical access constraints by frequency*

GLA will automatically rank each lexical access constraint according to frequency. The recognition of more frequently occurring words will be preferred.

As we saw in §3.4, a preference to recognize common words reduces the probability of confusion. I will now show that this desirable property of the recognition grammar is an automatic result of the Gradual Learning Algorithm, given the presence of a *LEX constraint for each lexical item.

To stay with our example of §3.5, suppose that an adult /Rat/ means |Rat| ‘rat’ 70% of the time, and |Rad| ‘wheel’ 30% of the time, but that the learner nevertheless has the ranking *LEX (|Rat| ‘rat’) >> *LEX (|Rad| ‘wheel’), which causes her to recognize the less common word |Rad| ‘wheel’ all of the time (if we ignore the phonology and the semantic context for a moment). We can now distinguish two cases. The first case applies 30% of the time: it is when the adult means |Rad| ‘wheel’. In this case, the learner will recognize correctly and see no reason to change her grammar. The second case, however, applies 70% of the time: the adult means |Rat| ‘rat’, but the learner comprehends |Rad| ‘wheel’. In this case, the learner will notice the discrepancy and change her grammar:

(43) *Learning as a result of misrecognition*

/Rat/ from Rat ‘rat’	*LEX (Rat ‘rat’)	*LEX (Rad ‘wheel’)
√ Rat ‘rat’	*!→	
*  * Rad ‘wheel’		←*

Since this situation is the more common case, the learner will, after a number of misrecognitions, end up with the correct ranking *LEX (|Rad| ‘wheel’) >> *LEX (|Rat| ‘rat’). If the final ranking difference would be large, however, the learner would recognize |Rat| ‘rat’ all of the time and have a misrecognition rate of 30%, causing the two constraints to approach each other again. As in §8.4, the final state is an equilibrium in which *LEX (|Rad| ‘wheel’) is ranked at such a distance above *LEX (|Rat| ‘rat’) that the learner will recognize |Rat| ‘rat’ 70% of the time and |Rad| ‘wheel’ 30% of the time. This is our second example of probability matching: the learner’s recognition probabilities will grow to equal the production probabilities of her language environment. In this example, the misrecognition rate will end up at $70\% \cdot 30\% + 30\% \cdot 70\% = 42\%$. This is higher than the error rate of the maximum-likelihood listener, which was 30%, but this cannot be helped (§15.3).

11.3 Ranking by semantic context


(34d) *Ranking of lexical access constraints by semantic context*

GLA will automatically determine the weight with which each semantic context contributes to the ranking of each *LEX constraint.

As we saw in §3.4, a preference to recognize words that are appropriate in the current semantic context reduces confusion. I will now show that this desirable property of the recognition grammar comes about automatically as a result of the GLA, given the existence of *LEX constraints for all lexical items in all semantic contexts.

To stay with our example, suppose that the learner has an anti-functional ranking, i.e., while a context of ‘turn’ ought to make her recognize /rat/ as |rad| ‘wheel’, she nevertheless has the ranking *LEX (|rad| ‘wheel’ / ‘turn’) >> *LEX (|rat| ‘rat’ / ‘turn’), which makes her recognize |rat| ‘rat’ all of the time. When the adult now says /rat/ in the context ‘turn’, and means |rad| ‘wheel’, the child wrongly comprehends |rat| ‘rat’, and if she notices the discrepancy, she will again take a learning step:

(44) *Learning as a result of misrecognition*

/rat/ context = ‘turn’ from rad ‘wheel’	*LEX (rad ‘wheel’ / ‘turn’)	*LEX (rat ‘rat’ / ‘turn’)
*  *	rat ‘rat’	←*
√	rad ‘wheel’	*!→

Eventually, the ranking of the two constraints will be reversed, and the learner will correctly recognize |rad| ‘wheel’ all of the time. If the adult does mean |rat| ‘rat’ 10 percent of the time, the learner will grow to match probabilities and rank *LEX (|rat| ‘rat’ / ‘turn’) at such a short distance above *LEX (|rad| ‘wheel’ / ‘turn’) that she will recognize |rad| ‘wheel’ only 90% of the time.

12. Faithfulness in recognition

12.1 Origin

(34e) *Origin of faithfulness constraints in recognition*

As soon as a perceptual category is created, the Language Acquisition Device supplies the recognition grammar with several faithfulness constraints (FAITH), which favour the similarity between the perceived form and the lexical form (with respect to the presence of features, their co-occurrence, and their sequencing).

For instance, when the learner creates the category /labial/ on the perceptual place tier, with nasality categories already in place, the recognition grammar will be supplied with a constraint *REPLACE (place: lab, cor / nasal), or */m/ → |n| for short, which favours the recognition of a word containing |m|, not |n|, every time that an /m/ will be perceived. The same for co-occurrence constraints like *DELETEPATH (place × nasal: labial × +), and faithfulness constraints for the surfacing of underlying temporal order and alignment.

12.2 Ranking

(34e) *Ranking of faithfulness constraints in recognition*

GLA will automatically rank these constraints according to the frequency of occurrence of the associated phonological feature values.

We proceed with the example of §3.2, and suppose that an intended |n| occurs three times as often as an intended |m| in the learner’s environment, and that there is a 9.6% probability for an /m/ perceived by the learner to have been intended by the adult as |n|, and a 4.2% probability for a perceived /n/ to have been intended as |m|. Suppose, then, that the lexicon contains the words |si:m| ‘seam’ and |si:n| ‘scene’, and the learner has the associated *LEX constraints ranked at equal height. Now suppose that the learner hears /si:m/. There are two cases in which she will take a learning step. The first case occurs if she recognizes /si:m/ as |si:m| ‘seam’ while the intended word was |si:n| ‘scene’:

(45) *Misrecognition*

/si:m/ intended: si:n ‘scene’	*REPLACE (lab, cor)	*LEX (si:m ‘seam’)	*LEX (si:n ‘scene’)
√ si:n ‘scene’	*!→		*→
** si:m ‘seam’		←*	

The probability that this happens, given the perception /si:m/, is $9.6\% \cdot p_d$, where p_d is the probability that *REPLACE outranks *LEX (|si:m| ‘seam’) during the evaluation. Because of noisy evaluation, there will be a second case, which occurs if the learner recognizes /si:m/ as |si:m| ‘scene’ while the intended word was |si:m| ‘seam’:

(46) *Misrecognition*

/si:m/ intended: si:m ‘seam’	*LEX (si:m ‘seam’)	*REPLACE (lab, cor)	*LEX (si:n ‘scene’)
** si:n ‘scene’		←*	←*
√ si:m ‘seam’	*!→		

The probability that this happens, given the perception /si:m/, is $90.4\% \cdot (1-p_d)$. The rerankings associated with the two misrecognitions will cancel each other out if $p_d=9.6\%$, i.e., if *REPLACE (lab, cor) outranks *LEX (|si:m| ‘seam’) by 1.8 standard deviations of the evaluation noise.

An analogous, equally probability-matching result will apply in the case of a perceived /si:n/: p_d will end up at 4.2%, which means that *REPLACE (cor, lab) will come to outrank *LEX (|si:n| ‘scene’) by 2.5 noise standard deviations. If other tableaux keep the two *LEX constraints at the same height, *REPLACE (cor, lab) will end up at 0.7 standard deviations above *REPLACE (lab, cor), which is the desirable ranking of (16). Unfortunately, quite a lot of complicating issues may disturb this picture: (1) the two tableaux above work only if the *LEX constraint for the unfaithful candidate is ranked low, which is impossible for both cases simultaneously; (2) other considerations, like semantic context, have to counteract the tendency of *LEX (|si:n| ‘scene’) to fall deeper than *LEX (|si:m| ‘seam’); (3) similar symmetries as in tableaux (45) and (46) will not work well in the cases of interactions of faithfulness with *LEXICALIZE and RECOGNIZE; (4) a part of the larger frequency of |n| may

stem from a larger token frequency for |n| words, which would lower the *LEX constraints containing a word with |n|, so that *LEX (|si:n| 'scene') must be expected to be ranked lower than *LEX (|si:m| 'seam'), thus causing a smaller difference in the *REPLACE rankings.

13. Faithfulness in production

13.1 Origin

(34f) *Origin of faithfulness constraints in production*

The Language Acquisition Device forces the use of the *same* faithfulness constraints, perhaps with the same ranking, in the production grammar as well as in the recognition grammar.


This means e.g. that it is equally bad for the speaker to pronounce an underlying |n| as /m/, as it is for the listener to recognize a form with /m/ as an underlying form with |n|. Hurford (1989) showed that evolutionarily, the best strategy is for the speaker to speak like other speakers speak, and for the listener to attend to the speaker's intentions, not to other listener's interpretations.

The frequency dependence noted under (34e) will thus lead to higher faithfulness constraints for less frequent feature values. This is realistic: the values [+nasal], [+round], and [labial] are less frequent than their counterparts [-nasal], [-round], and [coronal], and the first three have a tendency to beat their counterparts in cases of assimilation (the notion of privative features has its origin here).

13.2 Ranking

As long as the child is not yet capable of producing the feature values that she perceives, the relevant faithfulness constraints will raise in the grammar. Suppose, for instance, that the child has the feature /sibilant/, so that she perceives an adult [si:] as /si:/. Suppose further that she has stored the English word |si:| 'see' in her lexicon and that she wants to say this word. If she does not yet know how to produce sibilant noise, the outcome will be the articulation [ti:], which she will perceive as /ti:/:

(47) *Unfaithful production through lack of perceptuomotor knowledge*

si 'see' = adult /si:/	*DELETE (sibilant)
*  * [ti:] /ti:/	←*

I will explain why the candidate list does not contain any form that is perceived as /si:/. Any tableau in the production grammar must represent a part of the speaker's knowledge about the relation between articulation and perception, and about what structures violate what constraints. Therefore, all the articulatory candidates must be pronounceable forms. But the child does not yet know that for producing sibilant noise, she would have to execute a certain

tongue-grooving gesture, so no candidate [si:] (shortcut for something containing a tongue-grooving gesture) can appear in the candidate list.

Since she perceives the adult form /si:/ as different from her own form /ti:/, the learner will take action and promote *DELETE (sibilant) by a small step along the ranking scale. No number of these rerankings, however, will allow the learner to produce /si:/. Only when the relevant articulatory constraints enter the production grammar, the faithfulness constraints can hope for ultimate satisfaction (§14.2).

14. Articulatory effort

14.1 Origin

(34g) *Origin of articulatory constraints*

As soon as the child learns the perceptual result of an articulatory gesture, the Language Acquisition Device creates an articulatory constraint (*GESTURE) at the top of the production grammar.

We can assume that the learner will be able to home in on a desired perceptual result by playing about with her vocal apparatus. Thus, when the child learns that a certain tongue-grooving gesture, in combination with some already known laryngeal and velar gestures, produces the sibilant noise that she perceives as /s/, she will create a constraint *GESTURE (tongue: groove), and rank it high in her production grammar. This point of view is different from that by Smolensky (1996a), in which all structural constraints are ranked at the top of the grammar from the start. Instead, the presence of *GESTURE (tongue: groove) in the grammar already means that /s/ is *pronounceable*, i.e. the learner knows its perceptual consequences and will produce it during play or practice.


14.2 Ranking

(34g) *Ranking of articulatory constraints*

GLA will usually lower the gestural constraints, and automatically raise the corresponding faithfulness constraints, thus reranking the constraints in the direction of a grammar that produces a more adult-like output.

Initially, /s/ will not be realized faithfully in the communicative situation. The posited high ranking of *GESTURE (tongue: groove) refers to the effort during communication, in which this constraint has to compete with the faithfulness constraint *DELETE (sibilant):

(48) *Unfaithful production through lack of motor skill*

si 'see' = adult /si:/	*GESTURE (tongue: groove)	*DELETE (sibilant)
*  * [ti:] /ti:/		←*
√ [si:] /si:/	*!→	

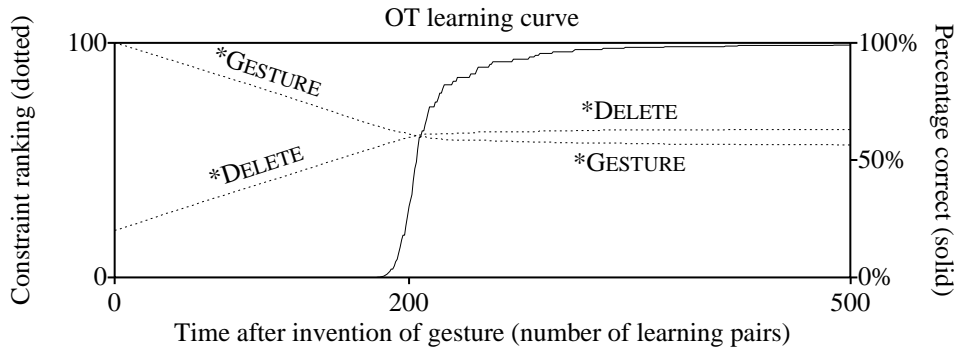


Fig. 5 Learning curve in the production grammar

Since the adult form /si:/ now shows up as the perceptual result of one of the learner's own candidate forms, the learner will not only have a faithfulness constraint to raise, but a gestural constraint to lower as well. After a number of rerankings, the child will be able to produce a faithful /si:/ in imitation (i.e. a less-linguistic situation with temporary lowering of gestural constraints and temporary raising of faithfulness, see Boersma 1998: 281), and after many rerankings, the constraints will have been reversed and the learner produces a faithful /si:/ most of the time. The existence of evaluation noise will ensure that this transition is not sudden; rather, the learner will show a smooth *learning curve*, with 50 percent adult-like /si:/ at the moment that the two constraints change order (Figure 5).

15. Discussion

15.1 Innateness

During the course of evolution, many functionally desirable properties have become innate. For phonology, these are:

(49) *Innate*

- a. The capability of perceptual categorization and its translation into PERCEIVE, *CATEG, and *WARP constraints.
- b. The capability of perceptual abstraction and its translation into OCP and LCC constraints.
- c. Motor skills in the speech tract (or for sign languages the upper half of the visible body) and their translation into *GESTURE.
- d. Initial high ranking of OCP in the perception grammar and *GESTURE in the production grammar.
- e. The capability of storing items in a large lexicon, and its translation into *LEXICALIZE.
- f. The capability of accessing items in the lexicon, and its translation into *LEX.
- g. Equal ranking of faithfulness in recognition and production (perhaps).
- h. Local ranking of faithfulness constraints by perceptual confusability, and local ranking of gestural constraints by articulatory effort.
- i. A gradual constraint-ranking learning algorithm that causes probability matching.

In a variable world, what is functionally desirable and what is not tends to change. To cope with this, evolution has endowed us with generic capabilities for learning the specific features of our actual environment. For phonology, these are:

(50) *Not innate*

- a. Language-specific perceptual features and feature values and their simultaneous and sequential combinations.
- b. Language-specific articulatory gestures and their simultaneous and sequential combinations.
- c. The capability for learning arbitrary language-specific generalizations (§15.2).

15.2 Generic constraint templates

Except for *LEXICALIZE and RECOGNIZE, all constraints proposed here are *templatic*: they contain parameters whose substance has to be filled in during acquisition. These parameters are the italicized parts in *REPLACE (*f*: *x*, *y* / *cond*), *GESTURE (*a*: *g* / *cond*), OCP (*f*: *x*; *a* | *m* | *b*) and so on. These parameters seem like an extra burden on the learner, if we compare these constraints with the more specific constraints proposed in earlier OT, like NOCODA and NASASSIM, which have their final form from the beginning and burden the learner with no more than the task to rank them in an order appropriate for the language they are learning.

But templatic constraints are needed anyway in phonology. McCarthy & Prince (1993) proposed a generic constraint family ALIGN (*constituent*₁, *edge*₁, *constituent*₂, *edge*₂), whose contents also have to be learned. Even more language-specific are morphological constraints, e.g. those that control verb forms in English such as “the past tense of *go* is *went*” and “the past tense is infinitive plus *-ed*”. For example, many kinds of specific morphologically conditioned output-output constraints have been proposed (Benua 1997). It seems, thus, that constraints that directly express the arbitrary facts of the language are needed in any case. If these can be learned, there is no argument against proposing the same for constraints that express faithfulness, abstraction, or effort.

15.3 Evolution of probability matching

In a stationary world, a human who always bases her choices on maximum-likelihood decisions would be best off. Why, then, is the probability-matching gradual learning algorithm only an approximation to this optimal situation? Well, the world is not stationary; the probabilities may change, and we can only detect this if we vary our behaviour somewhat. In a rapidly changing world, the best strategy is probably ‘tit-for-tat’ learning: always make the choice that would have been the best choice in the previous occasion. The tit-for-tat strategy obviously leads to probability-matching behaviour if the world *is* stationary. Thus, the gradual learning algorithm is an intermediate strategy: it shares with the maximum-likelihood strategy its criterion and its memory, and it shares with the tit-for-tat strategy its probability-matching property. This combination makes this algorithm perform reasonably well in a world where some probabilities are stationary, and some change rapidly. This functionally desirable behaviour, together with its straightforward implementation in terms of constraint reranking, may well have caused the successful evolution of moving multicellular creatures on the earth.

15.4 Subset problems

A recurring theme in the modelling of language acquisition is the ability of learning algorithms to cope with the *subset problem*: how does the learner, who can only learn from positive evidence, arrive at the most restrictive grammar of the language?

First, we may note that the Gradual Learning Algorithm solves the *overt subset problem*. The classical example is the English-speaking child who maintains two past tenses of the verb *go* (the verb of movement, not the board game), namely the rule-based form *goed* and the adult form *went*. Every time the learner would produce *goed*, but the adult says *went*, the learner will raise the constraint that favours *went*, and lower the *-ed* rule. Since the adult never says *goed*, the learner will end up saying *went* all of the time, unless other forms continue raising the *-ed* rule. The learner's own *goed* forms thus constitute the negative evidence necessary for remedying the overt subset problems. With forms that are variable in the adult grammar, like *dreamed* and *dreamt*, the learner will take her own *dreamt* form as negative evidence if the adult says *dreamed*, and she will consider her *dreamed* form incorrect if the adult says *dreamt*; these two effects balance each other, leading the child to ultimately copy the adult *dreamed-dreamt* ratio.

Second, the Gradual Learning Algorithm also handles the *covert subset problem* well. Under the maxim of Richness of the Base (Prince & Smolensky 1993: 191; Smolensky 1996b), it is the grammar, not the lexicon, that is responsible for enforcing restrictions on surface forms. Thus, if a language does not allow surface forms with codas, this has to be caused not by a lexicon that contains exclusively forms without codas, but by a grammar that would convert any (perhaps non-existent) lexical form with a coda to a form without. Generally, grammars will show this desirable property if structural constraints (which restrict possible output forms) are ranked as high as possible, and faithfulness constraints (which tend to maximize the number of possible output forms) are ranked as low as possible. In first approximation, therefore, this property is provided by starting from an initial state in which all structural constraints outrank all faithfulness constraints (Smolensky 1996a). But the learning algorithm EDCD by Tesar & Smolensky (1993, 1996, 1998) will quickly mess up this situation: if we start with all structural constraints in the first (highest) stratum and all faithfulness constraints in the second stratum, then as soon as a learning step requires the demotion of a certain structural constraint below a certain faithfulness constraint, that structural constraint will be demoted below *all* faithfulness constraints, immediately leading to overgeneration of surface forms under Richness of the Base. To tackle this situation, Hayes (1999) and Prince & Tesar (1999) propose complicated extensions to EDCD that cause some active enforcement of MARKEDNESS >> FAITHFULNESS rankings throughout the acquisition process. These extensions do not always lead to the most restrictive grammar, as these authors are well aware of. If we evaluate the GLA on this point, we see that its reaction to a learning datum that requires that a certain structural constraint be ranked below a certain faithfulness constraint, is to demote the structural constraint somewhat *and promote the faithfulness constraint* somewhat. This will typically lead to a selective rise of faithfulness constraints, causing much less overgeneration than with EDCD. The modest degree of overgeneration that typically remains, is comparable to that exhibited by real children (Boersma 2000b).

15.5 Consequences for phonological theory

Most theories of phonology cannot cope with non-innate substantive phonological material. Despite various proposals, nobody has yet given a satisfactory account for the feature [continuant] in feature geometry hierarchies (Clements 1985, Sagey 1986, McCarthy 1988). The so-called articulation-based feature hierarchies by Keyser & Stevens (1994) and Ladefoged (1997) are not that: both of them incorporate [place] and [nasal] under a [supralaryngeal] node, which only makes sense if we refer to the *perceptual* similarities of supralaryngeal articulations (namely, their formant patterns). The account given here solves the feature geometry problem: instead of forcing a single kind of cognitive features into a single hierarchy, we should just maintain two sets of features: perceptual features, with language-specific categorization and abstraction, and articulatory gestures, which consist of language-specific combinations of muscle movements. These two groups can only be put into shallow implicational hierarchies (Boersma 1998: 23).

15.6 Combining two of the three grammars

Smolensky (1996a) has observed that production and comprehension can be handled with the same kind of tableaux. Specifically, structural constraints, which evaluate surface forms, apply vacuously if the tableau is used for comprehension, because the surface form is identical for each candidate in comprehension. Smolensky's account has trouble, though, with phonological alternations, since in his model the comprehended form will always be equal to the surface form (Boersma 1999b). For this reason, we need to add lexical-access constraints to the grammars. Interestingly, this does not preclude the combination of the recognition and production grammars, since lexical-access constraints would apply vacuously during production, because they evaluate the underlying form, which is identical for each candidate in production. The possibility of combining the two grammars hinges, then, on the question whether the faithfulness constraints are ranked in the same order in the two grammars.

16. Conclusion

In this paper, I showed that it is plausible that the innate language acquisition device supplies the learner with a number of generic constraint templates in three grammars, and that the substantive content of these constraints is learned as a result of the language-specific development of perception and articulation. As a consequence, an adequate theory of autosegmental phonology should separate articulatory from perceptual representations and principles.

References

(ROA = Rutgers Optimality Archive, <http://rucss.rutgers.edu/roa.html>)

- Altmann, Gerry T.M. (ed., 1990). *Cognitive models of speech processing: psycholinguistic and computational perspectives*. Cambridge, Mass.: MIT Press.
- Bagley, W.C. (1900). 'The apperception of the spoken sentence: a study in the psychology of language.' *American Journal of Psychology* **12**: 80–130. [not seen]
- Beckman, Jill N. (1998). *Positional faithfulness*. Doctoral thesis, University of Massachusetts, Amherst.
- Behnke, Kay (1998). *The acquisition of phonetic categories in young infants: a self-organising artificial neural network approach*. Doctoral thesis, Universiteit Twente. [Max Planck Institute Series in Psycholinguistics **5**]
- Berko, Jean & Roger Brown (1960). 'Psycholinguistic research methods.' In Paul Mussen (ed.): *Handbook of research methods in child development*. New York: Wiley. 517–557.
- Benua, Laura (1997). *Transderivational identity: Phonological relations between words*. Doctoral thesis, University of Massachusetts, Amherst. [ROA **259**]
- Best, Catherine T., Gerald W. McRoberts & Nomathemba M. Sithole (1988). 'Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English speaking adults and infants.' *Journal of Experimental Psychology: Human Perception and Performance* **14**: 345–60.
- Boersma, Paul (1997). 'How we learn variation, optionality, and probability.' *Proceedings of the Institute of Phonetic Sciences* **21**: 43–58. University of Amsterdam. [<http://www.fon.hum.uva.nl/paul>]
- Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam. LOT International Series **11**. The Hague: Holland Academic Graphics. [<http://www.fon.hum.uva.nl/paul/diss/>]
- Boersma, Paul (1999a). *On the need for a separate perception grammar*. Ms. [ROA **358**]
- Boersma, Paul (1999b). 'Phonology-semantics interaction in OT, and its acquisition.' To appear in Robert Kirchner, Wolf Wikeley, & Joe Pater (eds.): *Papers in Experimental and Theoretical Linguistics*. Vol. 6. Edmonton: University of Alberta. [ROA **369**]
- Boersma, Paul (2000a). 'Nasal harmony in functional phonology.' Paper presented at HILP 4, Leyden, January 1999. [ROA]
- Boersma, Paul (2000b). 'A simple learning model with a realistic amount of subset problems.' Paper presented at the first North-American Phonology Conference, Montreal, April 2000. [ROA]
- Boersma, Paul (to appear). 'Learning a grammar in Functional Phonology.' In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.): *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford University Press. [included in ch. 14 of Boersma 1998]
- Boersma, Paul & Bruce Hayes (1999). *Empirical tests of the Gradual Learning Algorithm*. Ms. University of Amsterdam and UCLA. [ROA **348**]
- 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**]
- Braine, Martin D.S. (1976). 'Review of N.V. Smith, The acquisition of Phonology.' *Language*, **52**, 489–98.
- Carpenter, Gail A. & Stephen Grossberg (eds., 1991). *Pattern recognition by self-organizing neural networks*. Cambridge, Mass.: MIT Press.
- Cho, Taehong & Peter Ladefoged (1997). 'Variations and universals in VOT: evidence from 17 endangered languages.' *UCLA Working Papers in Phonetics* **95**: 18–40.
- Clements, G. Nick (1985). 'The geometry of phonological features.' *Phonology Yearbook* **2**: 225–252.
- Cole, Ronald A. & Brian Scott (1974). 'Toward a theory of speech perception.' *Psychological Review* **81**: 348–374.
- Cutler, Anne, Jacques Mehler, Dennis Norris & Juan Segui (1987). 'Phoneme identification and the lexicon.' *Cognitive Psychology* **19**: 141–177.
- Dupoux, Emmanuel & Jacques Mehler (1990). 'Monitoring the lexicon with normal and compressed speech: frequency effects and the prelexical code.' *Journal of Memory and Language* **29**: 316–335.
- Eggermont, Jos J. (1995). 'Representation of a voice onset continuum in primary auditory cortex of the cat.' *Journal of the Acoustical Society of America* **98**: 911–920.
- Eimas, Peter D. & John D. Corbit (1973). 'Selective adaptation of linguistic feature detectors.' *Cognitive Psychology* **4**: 99–109.
- Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk & James Vigorito (1971). 'Speech perception in infants.' *Science* **171**. 303–306.

- Elman, Jeffrey L. & James L. McClelland (1988). 'Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes.' *Journal of Memory and Language* **27**: 143–165.
- Foss, Donald J. & Michelle A. Blanck (1980). 'Identifying the speech codes.' *Cognitive Psychology* **12**: 1–31.
- Foss, Donald J. & David A. Swinney (1973). 'On the psychological reality of the phoneme: perception, identification and consciousness.' *Journal of Verbal Learning and Verbal Behavior* **12**: 246–257.
- Ganong, William F. III (1980). 'Phonetic categorization in auditory word perception.' *Journal of Experimental Psychology: Human Perception and Performance* **6**: 110–125.
- Garnes, Sara & Z.S. Bond (1976). 'The relationship between semantic expectation and acoustic information.' *Phonologica* **3**: 285–293.
- Gnanadesikan, Amalia (1995). *Markedness and faithfulness constraints in child phonology*. ROA **67**.
- Grossberg, Stephen (1976). 'Adaptive pattern classification and universal recoding: a parallel development and coding of neural feature detectors.' *Biological Cybernetics* **23**: 121–34.
- Hardcastle, William J. (1976). *Physiology of Speech Production. An Introduction for Speech Scientists*. Academic Press, London.
- Hayes, Bruce (1999). *Phonological acquisition in OT: the early stages*. ROA **327**.
- Hurford, James (1989). 'Biological evolution of the Saussurean sign as a component of the language acquisition device.' *Lingua* **77**: 187–222.
- Jun, Jongho (1995). 'Place assimilation as the result of conflicting perceptual and articulatory constraints.' *West Coast Conference on Formal Linguistics* **14**: 221–237.
- Jusczyk, Peter W., Derek Houston & Mara Goodman (1998). 'Speech perception during the first year.' In Alan Slater (ed.): *Perceptual development: visual, auditory, and speech perception in infancy*. Hove: Psychology Press. 357–388.
- Keyser, Samuel Jay & Kenneth N. Stevens (1994). 'Feature geometry and the vocal tract.' *Phonology* **11**: 207–236.
- Klatt, Dennis H. (1980). 'Speech perception: a model of acoustic-phonetic analysis and lexical access.' In Ronald A. Cole (ed.): *Perception and production of fluent speech*. Hillsdale: Erlbaum. 243–288.
- Kuhl, Patricia K. & James D. Miller (1978). 'Speech perception by the chinchilla: identification functions for synthetic VOT stimuli.' *Journal of the Acoustical Society of America* **63**: 905–17.
- Ladefoged, Peter (1997). 'Linguistic phonetic descriptions.' In William J. Hardcastle & John Laver (eds.): *The handbook of phonetic sciences*. Oxford & Malden, Mass.: Blackwell. 589–618.
- Lahiri, Aditi & William Marslen-Wilson (1991). 'The mental representation of lexical form: a phonological approach to the recognition lexicon.' *Cognition* **38**: 245–294.
- Liberman, Alvin M. & Ignatius G. Mattingly (1985). 'The motor theory of speech perception revised.' *Cognition* **21**: 1–36.
- Lisker, Leigh & Arthur S. Abramson (1964). 'A cross-language study of voicing.' *Word* **20**: 384–422.
- McCarthy, John (1988). 'Feature geometry and dependency: a review.' *Phonetica* **45**: 84–108.
- McCarthy, John & Alan Prince (1993). 'Generalized alignment.' In Geert Booij & Jaap van Marle (eds.): *Yearbook of Morphology 1993*. Dordrecht: Kluwer. 79–153. [ROA **7**]
- McClelland, James L. (1991). 'Stochastic interactive processes and the effect of context on perception.' *Cognitive Psychology* **23**: 1–44.
- McClelland, James L. & Jeffrey L. Elman (1986). 'The TRACE model of speech perception.' *Cognitive Psychology* **18**: 1–86.
- Macken, Marlys A. (1980). 'The child's lexical representation: The "puzzle-puddle-pickle" evidence.' *Journal of Linguistics* **16**: 1–17.
- MacNeilage, Peter F. (1997). 'Acquisition of speech.' In William J. Hardcastle and John Laver (eds.): *The handbook of phonetic sciences*. Oxford & Malden, Mass.: Blackwell. 303–32.
- McQueen, James M. (1991). 'The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity.' *Journal of Experimental Psychology: Human Perception and Performance* **17**: 433–443.
- 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.
- Mehler, Jacques, Jean Yves Dommergues, Uli Frauenfelder & Juan Segui (1981). 'The syllable's role in speech segmentation.' *Journal of Verbal Learning and Verbal Behavior* **20**: 298–305.
- Miller, Joanne L. & Emily R. Dexter (1988). 'Effects of speaking rate and lexical status on phonetic perception.' *Journal of Experimental Psychology: Human Perception and Performance* **14**: 369–378.
- Miller, Joanne L., Kerry Green & Trude M. Schermer (1984). 'On the distinction between prosodic and semantic factors in word identification.' *Perception & Psychophysics* **36**: 329–337.
- Norris, Dennis (1993). 'Bottom-up connectionist models of "interaction".' In G. Altmann & R. Shillcock (eds.): *Cognitive models of speech processing: the second Sperlonga meeting*. Hillsdale, NJ: Erlbaum. 211–234.
- Norris, Dennis (1994). 'Shortlist: a connectionist model of continuous speech recognition.' *Cognition* **52**: 189–234.

- Norris, Dennis & Anne Cutler (1988). 'The relative accessibility of phonemes and syllables.' *Perception & Psychophysics* **43**: 541–550.
- Padgett, Jaye (1995). 'Partial class behavior and nasal place assimilation.' *Proceedings of the Arizona Phonology Conference: Workshop on Features in Optimality Theory*, Coyote Working Papers, University of Arizona, Tucson. [ROA **113**]
- Pisoni, David B. & Paul A. Luce (1987). 'Acoustic-phonetic representations in word recognition.' *Cognition* **25**: 21–52.
- Polka, Linda & Janet F. Werker (1994). 'Developmental changes in perception of non-native vowel contrasts.' *Journal of Experimental Psychology: Human Perception and Performance* **20**: 421–435.
- Pols, Louis C.W. (1983). 'Three-mode principal component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation.' *Speech Communication* **2**: 275–293.
- Powers, William T. (1973). *Behavior: The control of perception*. Chicago: Aldine.
- Prince, Alan & Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report TR-2, Rutgers University Center for Cognitive Science.
- Prince, Alan & Bruce Tesar (1999). *Learning phonotactic distributions*. Technical Report TR-54, Rutgers University Center for Cognitive Science. [ROA **353**]
- Quinn, Paul C., Peter D. Eimas & S.L. Rosenkrantz (1993). 'Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants.' *Perception* **22**: 463–475.
- Sagey, Elizabeth (1986). *The representation of features and relations in nonlinear phonology*. Doctoral thesis, MIT, Cambridge.
- Samuel, Arthur G. (1981). 'The role of bottom-up confirmation in the phonemic-restoration illusion.' *Journal of Experimental Psychology: Human Perception and Performance* **7**: 1124–1131.
- Samuel, Arthur G. (1990). 'Using perceptual-restoration effects to explore the architecture of perception.' In Altmann (ed.), 295–314.
- 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]
- Segui, Juan, Uli Frauenfelder & Jacques Mehler (1981). 'Phoneme monitoring, syllable monitoring and lexical access.' *British Journal of Psychology* **72**: 471–477.
- Smith, Neilson V. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.
- Smolensky, Paul (1996a). 'On the comprehension/production dilemma in child language.' *Linguistic Inquiry* **27**: 720–731.
- Smolensky, Paul (1996b). *The initial state and 'richness of the base' in Optimality Theory*. Technical Report **96-4**, Department of Cognitive Science, Johns Hopkins University, Baltimore. [ROA **154**]
- Stevens, Kenneth N. (1998). *Acoustic Phonetics*. Cambridge, Mass. & London: MIT Press.
- Swoboda, P., P.A. Morse & L.A. Leavitt (1976). 'Continuous vowel discrimination in normal and at-risk infants.' *Child Development* **49**: 332–339.
- Tesar, Bruce & Paul Smolensky (1993). *The learnability of Optimality Theory: an algorithm and some basic complexity results*. Ms. Department of Computer Science & Institute of Cognitive Science, University of Colorado at Boulder. [ROA **2**]
- Tesar, Bruce & Paul Smolensky (1996). *Learnability in Optimality Theory (long version)*. Technical Report **96-3**, Department of Cognitive Science, Johns Hopkins University, Baltimore. [ROA **156**]
- Tesar, Bruce & Paul Smolensky (1998). 'Learnability in Optimality Theory.' *Linguistic Inquiry* **29**: 229–268.
- Vihman, Marilyn May (1982). 'A note on children's lexical representations.' *Journal of Child Language*, **9**, 249–53.
- Vihman, Marilyn May (1996). *Phonological development. The origins of language in the child*. Cambridge, Mass., and Oxford: Blackwell.
- Werker, Janet F. & R.C. Tees (1984). 'Cross-language speech perception: evidence for perceptual reorganization during the first year of life.' *Infant Behavior and Development* **7**: 49–63.
- Wickelgren, Wayne A. (1969). 'Context-sensitive coding, associative memory, and serial order in (speech) behavior.' *Psychological Review* **76**: 1–15.
- Zoll, Cheryl S. (1998). *Positional asymmetries and licensing*. ROA **282**.