Modelling the formation of phonotactic restrictions across the mental lexicon

Hamann, S.; Apoussidou, D.; Boersma, P.

Publication date
2009

Document Version
Final published version

Published in
Chicago Linguistic Society. Papers from the Regional Meetings

Modelling the formation of phonotactic restrictions across the mental lexicon*

Silke Hamann[a], Diana Apoussidou[b] & Paul Boersma[b]
[a]Heinrich-Heine-Universität Düsseldorf and [b]Universiteit van Amsterdam

1 Introduction

Experimental data shows that adult learners of an artificial language with a phonotactic restriction learned this restriction better when being trained on word *types* (e.g. when they were presented with 80 different words twice each) than when being trained on word *tokens* (e.g. when presented with 40 different words four times each) (Hamann & Ernestus submitted). These findings support Pierrehumbert's (2003) observation that phonotactic co-occurrence restrictions are formed across lexical entries, since only lexical levels of representation can be sensitive to type frequencies.

Current models of language learnability (e.g. Hayes & Wilson 2008) can replicate the distributions of phonotactic restrictions in the input, but not the qualitative distinction between the different sources of these frequencies. That is, they predict the same change in grammar when a single word with a phonotactic pattern is presented twice in comparison to when two words with the same phonotactic pattern are presented once each. What is lacking in these models is a level of lexical representation that enables the learner to distinguish between type and token frequencies. We present a computational model that can explain the type- vs. token learning effects, namely Bidirectional Phonology and Phonetics (Boersma 2007) with an additional semantic level (Apoussidou 2007).

The present paper is structured as follows. In §2 we give an overview of the experimental findings in the study by Hamann & Ernestus (submitted). In §3 we introduce the linguistic model we employ to account for the experimental findings. Section 4 illustrates the learning algorithm and the learning steps that our learners make. In §5 we conclude and address topics for further research.

2 The data

The data stems from two psycholinguistic experiments by Hamann & Ernestus (submitted), where a total of 36 adult Dutch native speakers had to learn an artificial language. The language consisted of bisyllabic words of the form given in (1) with stress on the first syllable.

(1) VC(j)V

* We would like to thank the audience at the 16th Manchester Phonology Meeting and at the 45th Meeting of the Chicago Linguistics Society for helpful comments and questions.

The segments were chosen from the existing Dutch vowels {i e a u} and the existing Dutch consonants {b p d t k s f}. The palatal glide [j] occurred before the front vowels [i e] but not before the non-front vowels [a u]. A phonotactic restriction like this does not exist in Dutch.[1]

Examples of words allowed in the artificial language are given in (2a) and (b).

(2) a. [afji, ikje, upji, etje]
    b. [uta, isu, eba, adu]

Participants first underwent a training phase, in which they heard a number of word forms from the artificial language, each of which was presented together with a picture depicting its meaning. The participants were asked to listen to these words and to look at the pictures. In the subsequent test phase they heard new words and had to judge whether these were possible words in the language they just heard. This test phase consisted of 70 words, 24 of which violated the phonotactic restriction, as in the examples in (3).[2]

(3) a. [ifju, usja, edja, abju]
    b. [uke, epi, ate, ifi]

All words used in the training and test conditions were produced in a natural way by the same female Dutch native speaker.

The number of words in the training phase varied according to three conditions, summarized in (4).

(4) Experiment 1	(total of 80 words):	40 words repeated 2 times
Experiment 2a	(total of 160 words):	40 words repeated 4 times
Experiment 2b	(total of 160 words):	80 words repeated 2 times

The data of the test phase were analyzed by means of multi-level logistic regression analyses with word and participant as crossed random factors (see e.g. Jaeger 2008). The "percentage correct" shown in Fig. 1 was computed as the grand sum of the acceptance of correct words and the rejection of incorrect words under each condition.

[1] Dutch allows sequences of obstruent plus /j/ plus full vowel at the end of a word, though these sequences are restricted to a very limited number of words (e.g. *Atjeh* [atje] proper name).

[2] The first 12 items were used to familiarize the participants with the task. They were mostly repetitions of words that occurred in the training phase, and were not included in the analysis. Of the remaining 58 words, 12 included words with sonorants such as [anji, ime, elu, uŋa] to test whether the learners had acquired the inventory restriction, i.e. the set of consonants allowed in the artificial language.

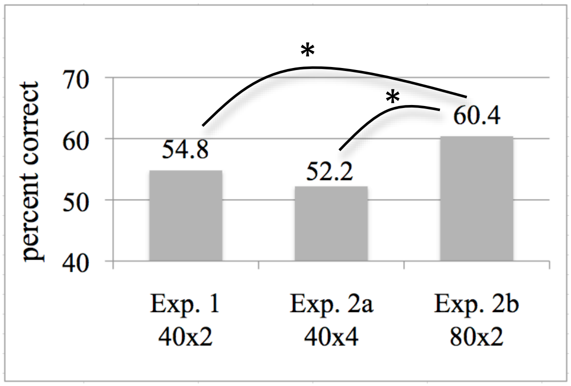

Figure 1: Overall percentage correct in the test phases of Experiment 1 and the two conditions of Experiment 2. * indicates a *p*-value smaller than 0.05.

As we can see in Fig. 1, 80 words occurring twice resulted in a better performance than 40 words occurring two or four times (no significant difference in performance was detected between 40 words occurring two or four times). The phonotactic restriction is hence formed more easily if it is made on the basis of more word types, even with the same total number of presentations (160).

Since any differentiation between word types and tokens can only be made with the help of the mental lexicon, the results provide evidence that phonotactic generalizations are made across lexical entries, in accordance with Pierrehumbert's (2003) observation that "[p]honotactic generalizations are made over all words in the lexicon" (p. 8).

3 The linguistic model

The behaviour of the participants in Hamann & Ernestus' task seems to pose a problem for on-line learning models. Simple connectionist and similar phonological models of processing (TRACE: McClelland & Elman 1986; Harmonic Grammar: Smolensky & Legendre 2006; Stochastic Optimality Theory: Boersma & Hayes 2001) update their connection weights or constraint rankings at every incoming piece of data. Superficially, one could be inclined to predict that such models are sensitive to token frequency rather than to type frequency.

We handle this alleged problem by modeling the acquisition of the phonotactic restrictions with a model that combines the phonological and phonetic levels of a language with a lexical level of meaning. The lexical level is necessary to allow a distinction between type and token frequency. Phonotactic learning is restricted to the occurrence of new word types. This is implemented in the following way. The first time that a word form (with meaning) occurs, the virtual learner creates a strong connection (in her lexicon) between the given underlying form and the given meaning ('one-shot learning'), and demotes a constraint punishing this connection. By virtue of the general properties of the learning algorithm, this demotion comes with the concomitant demotion of the relevant phonotactic constraint. Further occurrences of the same word do not trigger any creations of connections, hence cannot change the strengths of the phonotactic constraints; however, occurrences of new similar words do trigger the demotion of a lexical constraint and hence the concomitant demotion of the phonotactic constraint.

The model we employ in the present study is the interactive multiple-level bidirectional model of phonology and phonetics (BiPhon; Boersma 2007), extended with a semantic-morphemic level of representation from Apoussidou (2007). This model distinguishes (at least) four levels of representation, given in the middle of Fig. 2.[3]

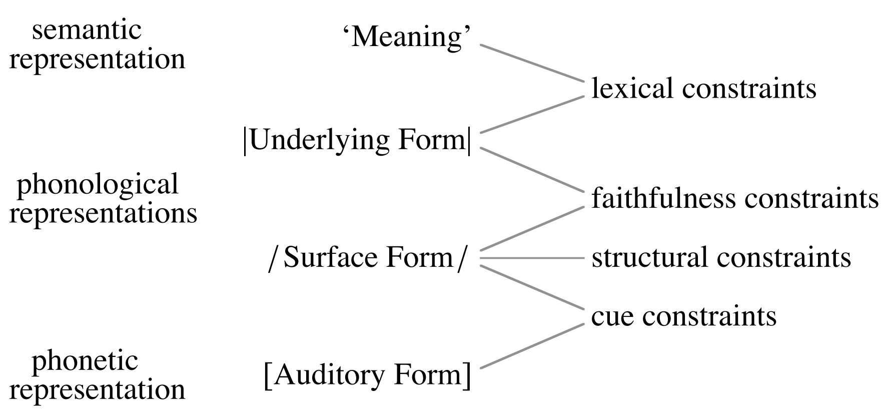

Figure 2: Representations and connections in the bidirectional model of phonology and phonetics.

A crucial property of this model is that underlying phonological forms are not ‘stored’ in the lexicon but ‘emerge’ every time a speaker has to compute the pronunciation of a morpheme.

Connections between the representations are formalized as constraints, and so are the restrictions on representations, as shown on the right side of figure 2. The following four types of constraints are relevant for our simulations:[4]

(5) a. *Lexical constraints*:
“Do not connect the meaning ‘*x*’ with the underlying form |*y*|”,
e.g. *‘house’|ata|.

---

[3] The ‘Meaning’ representation can be further split up into morphemes and context, and the auditory form can be supplemented by an articulatory form, see Boersma (to appear).

[4] The surface and the underlying form are abstract representations, and the notation used here is shorthand for the respective feature combinations. The auditory form, on the other hand, is a concrete, detailed representation, and the notation stands for auditory cues such as first and second formant, duration, and so on.

b. *Faithfulness constraints*:
"Do not connect the underlying form |*x*| with the surface form /*y*/",
e.g. *|ata|/atja/.

c. *Structural constraints*:
"Do not realize the surface form /*x*/",
e.g. */ja/.

d. *Cue constraints*:
"Do not connect the surface form /*x*/ with the auditory form [*y*]",
e.g. */atja/[atje].

Faithfulness (5b) and structural constraints (5c) have been familiar to phonologists since Prince & Smolensky (1993), while lexical (5a) and cue constraints (5d) were introduced by the present multilevel-OT framework (lexical constraints in Boersma 2001, cue constraints in Boersma 1998: 241).

4 Simulated learning

The language we used in our simulations has in total 20 words; all of them have the structure in (1) and thus follow the phonotactic restriction employed by Hamann & Ernestus.

We simulated hundred virtual learners with two constraint-based learning procedures available in the Praat program (Boersma & Weenink 2009), namely Stochastic Optimality Theory (Boersma 1998) and Noisy Harmonic Grammar (Boersma & Escudero 2008). Both use the error-driven Gradual Learning Algorithm (GLA) for OT (Boersma 1997, Boersma & Hayes 2001) or HG (Soderstrom, Mathis & Smolensky 2006, Boersma & Pater 2008). The GLA causes a change in the ranking (for OT) or weighting (for HG) of constraints if the candidate that the learner would choose (either in production or perception or in both processing directions) is not the one observed in the learner's input. In such cases, the constraints that prefer the input are promoted or get more weight, and those that prefer the wrongly winning candidate are demoted or get less weight.

We will focus in the following on OT learning, the differences with the HG learning procedure are summarized in §4.4 below.

4.1 The initial grammar

Each of the fifty OT learners started with an initial grammar that had 20 lexical constraints, one for every word of the language, ranked very high (with a ranking value of 1000 on the ranking scale); see the first three rows in (6). This high-ranking of the lexical constraints indicates that the learners did not have any connections between meanings and underlying forms, i.e. no lexical entries, for the new language when they started learning. In addition, we used the general lexical constraint *' '|*x*|, which requires that an underlying form has a meaning assigned to it, i.e. that lexical entries must include a meaning component. Like the other lexical constraints, this constraint was assigned an initial ranking value of 1000.

The initial grammar furthermore contained phonotactic constraints against both the allowed and the disallowed glide-vowel sequences of the artificial language; they are enumerated in the last eight rows in (6), where the underscore stands for the absence of the palatal glide. These phonotactic constraints were ranked lowest in the grammar, with an initial ranking value of 100.

The grammars also contained an aggregate faithfulness constraint FAITH, punishing all non-faithful mappings between surface and underlying form, and an aggregate cue constraint CUE, punishing all divergences between auditory cues and their surface forms. These two constraints were used instead of numerous detailed cue and faithfulness constraints because the focus of the study was the learning of the lexical and the phonotactic constraints. The initial constraints and their rankings looked as follows:

(6)

Constraint	*ranking value*	*plasticity*
*|ata|‘house’	1000	2000
*|ifa|‘chair’	1000	2000
[…]		
*‘ ’|*x*|	1000	1
FAITH	500	1
CUE	300	1
*/ji/	100	1
*/je/	100	1
*/ja/	100	1
*/ju/	100	1
*/_i/	100	1
*/_e/	100	1
*/_a/	100	1
*/_u/	100	1

The plasticity values in the last column of (6) indicate the steps that the constraints are moved up or down the ranking scale in case of learning. The plasticities of the lexical constraints are large because we assume that the lexical constraints are learned fast (one-shot learning). This is based on the observation that real language learners usually know a word once they have heard it and can associate a meaning with it. All other constraints are moved more slowly with a plasticity value of one.

In addition to the listing of the constraints and their rankings, the initial grammar included a list of possible input forms and the constraint violations they incur. For every word in the simulated language, there are two possible input forms: a meaning and an auditory form. The meaning is the input to the production process, and the auditory form the input to the comprehension process (BiPhon takes into account both directions of processing).

Candidates are always quadruplets of meaning, underlying form, surface form and auditory form, because we are operating with a multi-level grammar that handles the connections between all four levels. We reduced the possible candidate list to four candidates for each auditory input, as exemplified in (7) for the input [ata]. These four candidates illustrate all possible kinds of constraint violations.[5] Only the first is also a candidate for the input of the meaning 'house' in the production process.

(7)

	1000	1000	1000	500	300	100	100	100	100	100	100	100	100
[ata]	*|*x*| ' '	*|ata| 'house'	*|ifa| 'chair'	FAITH	CUE	*ji	*je	*ja	*ju	*_i	*_e	*_a	*_u
'house' |ata| /ata/ [ata]		*!										*	
' ' |ata| /ata/ [ata]	*!											*	
' ' | | /ata/ [ata]				*!								*	
☞ ' ' | | / / [ata]					*								

The first candidate combines most faithfully a representation on every level, but violates the respective lexical constraint. The second candidate has an underlying representation but no semantic form connected to it, violating *' '|*x*|. The third candidate has a surface representation but no underlying one, violating FAITH, and the fourth candidate is an auditory form with no surface representation connected to it, violating CUE.

4.2 Learning

To trigger learning, the virtual learners were fed with input pairs consisting of meaning and overt form like 'house' [ata], similar to the training conditions in Hamann & Ernestus. We constructed two training sets to test the effects of word types versus word tokens on phonotactic learning that were observed by Hamann & Ernestus. Half of the learners were presented with twenty different input pairs once (in randomized order). This training type is called *type training* in the following. The second half was presented with ten input pairs twice (in randomized order), called *token training*. The list of all input pairs for both training types is given in the appendix.

The learning tableau in (8) illustrates the learning process, with the example 'house' [ata]. This tableau shows three winners: the form marked "√" is the most optimal quadruplet in which the auditory form is [ata] and the meaning is 'house',

[5] Further candidates all involve either a violation of a lexical constraint, *' '|*x*|, CUE, or FAITH, or several of them, and do not influence learning. This also holds for the candidate 'house' |ata| /atja/ [atja], which violates FAITH and a structural constraint against the non-occurring sequence /ja/, and the candidate 'house' |ata| /ata/ [atja], which violates CUE and the structural constraint */_a/.

i.e. the learner can consider this the ‘correct’ form because it takes into account all available input data (both sound and meaning); the form marked “☜” is the winner of the comprehension process, i.e. the most optimal of all candidates in which the auditory form is [ata]; and the form marked “☞” is the winner of the production process, i.e. the most optimal of all candidates in which the meaning is ‘house’. In our restricted tableaus, the production winner always happens to equal the ‘correct’ form. Error-driven learning, then, can only be triggered if the comprehension winner differs from the correct form, as it does in (8). In such a case, the learner will move the constraints to arrive at a grammar with only a single optimal candidate, i.e. a grammar that does not produce errors. The general learning procedure, as the arrows in (8) illustrate, consists of a demotion of both constraints violated by the ‘correct’ quadruplet (a lexical constraint and a phonotactic constraint), and a promotion of the one constraint that is violated by the comprehension winner (the CUE constraint).

(8)

	1000	1000	1000	500	300	100	100	100	100	100	100	100	100
‘house’ [ata]	*|*x*| ‘ ’	*|ata| ‘house’	*|ifa| ‘chair’	FAITH	CUE	*ji	*je	*ja	*ju	*_i	*_e	*_a	*_u
√ ☞ ‘house’ |ata| /ata/ [ata]		*→										*→	
‘ ’ |ata| /ata/ [ata]	*											*	
‘ ’ | | /ata/ [ata]				*								*	
☜ ‘ ’ | | / / [ata]					←*								

After the demotion and promotion of the constraints, the grammar looks as in (9). The lexical constraint *‘house’|ata|, has been demoted by 2000, i.e. all the way to the bottom of the ranking scale; the constraint against the observed sequence /_a/ ends up below the competing constraint against the non-occurring sequence /ja/; and CUE is slightly promoted.

(9)

	1000	1000	500	301	100	100	100	100	100	100	100	99	–1000
‘house’ [ata]	*|*x*| ‘ ’	*|ifa| ‘chair’	FAITH	CUE	*ji	*je	*ja	*ju	*_i	*_e	*_u	*_a	*|ata| ‘house’
√ ☞ ☜ ‘house’ |ata| /ata/ [ata]												*	*
‘ ’ |ata| /ata/ [ata]	*											*	
‘ ’ | | /ata/ [ata]			*									*	
‘ ’ | | / / [ata]				*									

The resulting new grammar is such that the next time the input [ata] ‘house’ arrives, the comprehension winner will equal the ‘correct’ form. Consequently, a second occurrence of the same word type cannot trigger any learning, because it does not produce an error. Training with repetitions of the same words, as in our *token training*, thus does not help learning.

What happens if the learner encounters a different word type instead, as in *type training*, is shown in (10). The new word type violates a high-ranked lexical constraint, which is the reason for the difference between the correct winner and the comprehension winner.[6]

(10)

<table>
<tr><td></td><td>1000</td><td>1000</td><td>500</td><td>301</td><td>100</td><td>100</td><td>100</td><td>100</td><td>100</td><td>100</td><td>100</td><td>99</td><td>–1000</td></tr>
<tr><td>‘chair’ [ifa]</td><td>*|x| ‘ ’</td><td>*|ifa| ‘chair’</td><td>FAITH</td><td>CUE</td><td>*ji</td><td>*je</td><td>*ja</td><td>*ju</td><td>*_i</td><td>*_e</td><td>*_u</td><td>*_a</td><td>*|ata| ‘house’</td></tr>
<tr><td>√ ☞ ‘chair’ |ifa| /ifa/ [ifa]</td><td></td><td>*→</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*→</td><td></td></tr>
<tr><td>‘ ’ |ifa| /ifa/ [ifa]</td><td>*</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*</td><td></td></tr>
<tr><td>‘ ’ | | /ifa/ [ifa]</td><td></td><td></td><td>*</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*</td><td></td></tr>
<tr><td>☜ ‘ ’ | | / / [ifa]</td><td></td><td></td><td></td><td>←*</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

This difference between the correct form and the comprehension winner in (10) results in a further demotion of the phonotactic constraint, as indicated by the arrow for */_a/. The result of this learning step is shown in (11).

(11)

<table>
<tr><td></td><td>1000</td><td>500</td><td>302</td><td>100</td><td>100</td><td>100</td><td>100</td><td>100</td><td>100</td><td>100</td><td>98</td><td>–1000</td><td>–1000</td></tr>
<tr><td>‘chair’ [ifa]</td><td>*|x| ‘ ’</td><td>FAITH</td><td>CUE</td><td>*ji</td><td>*je</td><td>*ja</td><td>*ju</td><td>*_i</td><td>*_e</td><td>*_u</td><td>*_a</td><td>*|ata| ‘house’</td><td>*|ifa| ‘chair’</td></tr>
<tr><td>√ ☞☜ ‘chair’ |ifa| /ifa/ [ifa]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*</td><td></td><td>*</td></tr>
<tr><td>‘ ’ |ifa| /ifa/ [ifa]</td><td>*</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*</td><td></td><td></td></tr>
<tr><td>‘ ’ | | /ifa/ [ifa]</td><td></td><td>*</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*</td><td></td><td></td></tr>
<tr><td>‘ ’ | | / / [ifa]</td><td></td><td></td><td>*</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

We conclude that *type training* does help learning. Together, the single demotion of */_a/ in the token training in (8) and (9) and the multiple demotions of */_a/ in the type training in (8) and (10) show that our OT account has successfully modelled the differential effect of type versus token frequency on the acquisition of a phonotactic constraint.

4.3 The final grammar

The final stages for the grammars of the 25 learners with type training and the 25 learners with token training look as follows:

---

[6] With the tableaus in (8) and (10) we can illustrate why CUE must be lower ranked than FAITH: all candidates apart from the last one violate the phonotactic constraint */_a/ because they have a surface form containing this sequence. Only a candidate without a surface structure that wins in production or comprehension can cause a re-ranking of this phonotactic constraint. If CUE were ranked above FAITH, the third candidate would be the winner of the comprehension process, which would not cause a demotion of */_a/.

(12) a. Type training

Constraint	ranking value
*‘ ’|x|	1000
FAITH	500
CUE	320
*/ji/	100
*/je/	100
*/ja/	100
*/ju/	100
*/_i/	95
*/_e/	95
*/_a/	95
*/_u/	95
20 lexical constraints	–1000

b. Token training

Constraint	ranking value
10 lexical constraints	1000
*‘ ’|x|	1000
FAITH	500
CUE	310
*/ji/	100
*/je/	100
*/ja/	100
*/ju/	100
*/_i/	98
*/_e/	98
*/_a/	97
*/_u/	97
10 lexical constraints	–1000

We can see that the final grammars differ in three respects. First, the learners with type training have all lexical constraints demoted, while those with token training demoted only ten lexical constraints; this is simply due to the fact that learners with token training encountered only ten different word types whereas the learners with type training received twenty different word types. The second difference is the ranking value of CUE, which ends up higher in the type training because it is raised with every new word type (recall tableaus (8) and (10)). Lastly and most importantly, the ranking values of the phonotactic constraints against the occurring sequences are different. In the grammar of the token learners these four constraints have the values 97 or 98, in the type learner’s grammar they are all at 95, i.e. they have been demoted twice as far as those of the token learners, directly reflecting the fact that the type learners heard twice as many types as the token learners.

4.4 Any differences for Harmonic-Grammar learning?

The Harmonic-Grammar learning procedure differs from the OT procedure in that constraints have weights instead of rankings. When we use Noisy HG in our simulations, the resulting weights are equal to the ranking values in (12). To illustrate this point, the OT learning tableau from (8) is given in (13) as an HG learning tableau (without the lexical constraint *‘chair’|ifa|, for lack of space). The numbers above the constraint names are now the weights of the constraints, not their ranking values. The last column gives the harmonies of the candidates.

(13)

	1000	1000	500	300	100	100	100	100	100	100	100	100	
‘house’ [ata]	*|*x*| ‘ ’	*|ata| ‘house’	FAITH	CUE	*ji	*je	*ja	*ju	*_i	*_e	*_a	*_u	
√☞‘house’ |ata| /ata/ [ata]		–1→									–1→		–1100
‘ ’ |ata| /ata/ [ata]	–1										–1		–1100
‘ ’ | | /ata/ [ata]			–1								–1		–600
☜ ‘ ’ | | / / [ata]				← -1									–300

Errors like those in tableau (13) lead the learner to update her constraint weights. After simulating learning, the final grammar is the same as in OT case of §4.3. Noisy HG learners thus turn out to have the same type-token sensitivity as the Stochastic OT learners.

5 Conclusion

The goal of this paper was to model the experimental data found by Hamann & Ernestus (submitted) who showed that adults learnt a phonotactic restriction better when presented with many word types than when presented with a smaller set of word types that were repeated. The combined result of virtually learning the connections between four levels of representation (meaning, underlying form, surface form and auditory form) with a single learning algorithm is a sensitivity to type frequency: virtual learners trained on 20 word types (presented once) ended up with stronger phonotactic constraints than those trained on 10 word types (presented twice), predicting that the former group is better at generalizing the learned phonotactic restrictions.

Our proposed model is much simpler than previous generative accounts of phonotactic learning (such as Tesar & Smolensky 1998, Prince & Tesar 2004, Hayes & Wilson 2008), yet makes the same kind of generalizations. However, unlike these previous models, it does this in a way that is compatible with psycholinguistic findings (Hamann & Ernestus submitted).

As illustrated in §4.2, our model comes with a number of assumptions, the most important one being that the different types of constraints in the initial grammar have different ranking values and partly also different plasticity values. First, we assume that all lexical constraints are initially high ranked and have a high plasticity value. Our motivation for this assumption is the fact that adult learners do not have any a priori connections between meaning and form for a new language, but once they encounter a new word (a form-meaning pair) in this language, they create such a connection, which triggers a strong demotion of the respective constraint (one-shot learning, implemented with a high plasticity value for lexical constraints).

Second, we have to postulate that FAITH is ranked above CUE (see footnote 6 for why this is essential for our analysis). Also, CUE must be ranked above the structural constraints in our initial grammars, otherwise the wrong candidate would win in the comprehension direction in tableau (9). These postulations could be motivated by the fact that adult learners have an initial ranking of Faithfulness and Cue constraints based on their native language.

Lastly, we have to assume that the learning tableaus, in order to trigger learning, include a candidate that violates CUE but not the structural constraints. Given that OT approaches postulate a function GEN that generates all possible candidates, this last assumption is unproblematic.

All our assumptions are made on the basis that the learner is an L2 adult, i.e. someone who went through the initial acquisition stages for her native language. We do not propose that the initial rankings in (6) hold for infant acquisition, nor do we propose that one-shot lexical learning applies to infants. Future research is necessary to elaborate and account for the difference between infant and adult phonotactic learning.

References

Apoussidou, D. 2007. *The Learnability of Metrical Phonology*. Ph.D. dissertation, Univ. of Amsterdam.

Boersma, P. 1997. How we learn variation, optionality, and probability. *Proc. Institute of Phonetic Sciences, Univ. of Amsterdam* 21.43–58.

Boersma, P. 1998. *Functional Phonology: Formalizing the Interaction between Articulatory and Perceptual Drives*. Ph.D. dissertation, Univ. of Amsterdam.

Boersma, P. 2001. Phonology-semantics interaction in OT, and its acquisition. In *Papers in Experimental and Theoretical Linguistics*, ed. by R. Kirchner, W. Wikeley & J. Pater, 24–35. Edmonton: University of Alberta.

Boersma, P. 2007. Some listener-oriented accounts of *h-aspiré* in French. *Lingua* 117.1989–2054.

Boersma, P. to appear. A programme for bidirectional phonology and phonetics and their acquisition and evolution. In *Bidirectional Optimality Theory*, ed. by A. Benz & J. Mattausch.

Boersma, P. & P. Escudero. 2008. Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. In *Contrast in Phonology: Theory, Perception, Acquisition*, ed. by P. Avery, B. E. Dresher & K. Rice, 271–301. Berlin: Mouton de Gruyter.

Boersma, P. & B. Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32.45–86.

Boersma, P. & J. Pater. 2008. Convergence properties of a gradual learning algorithm for Harmonic Grammar. *Rutgers Optimality Archive* 970. http://roa.rutgers.edu

Boersma, P. & D. Weenink. 2009. Praat: doing phonetics by computer (Version 5.1.04). http://www.praat.org/. Retrieved 08 Apr 09.

Hamann, S. & M. Ernestus. submitted. Adult learning of inventory and phonotactic restrictions: the role of type and token frequency.

Hayes, B. & C. Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.

Jaeger, F. T. 2008. Categorical data analysis: Away from ANOVAs transformation or not and towards logit mixed models. *J. Memory and Language* 59.434–446.

McClelland, J. & J. L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18.1–86.

Pierrehumbert, J. B. 2003. Probablilistic phonology: Discrimination and robustness. In *Probability Theory in Linguistics*, ed. by R. Bod, J. Hay & S. Jannedy, 177–228. Cambridge, Mass.: MIT Press.

Prince, A. & P. Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Published 2004 by Blackwell, London.

Prince, A. & B. Tesar. 2004. Learning phonotactic distributions. In *Constraints in Phonological Acquisition*, ed. by R. Kager, J. Pater and W. Zonneveld, 245–291. Cambridge: Cambridge University Press.

Smolensky, P. & G. Legendre. 2006. *The Harmonic Mind*. Cambridge Mass.: MIT Press.

Soderstrom, M., D. Mathis & P. Smolensky. 2006. Abstract genomic encoding of Universal Grammar in Optimality Theory. In Smolensky & Legendre, 967–1037.

Tesar, B. & P. Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29.229–268.

Appendix: The two training sets

Type training
(all words presented once):

‘house’ [ata]
‘car’ [uba]
‘table’ [eda]
‘chair’ [ifa]
‘lamp’ [asa]
‘sofa’ [apu]
‘pot’ [ufu]
‘cat’ [etu]
‘dog’ [isu]
‘flower’ [eku]
‘shoes’ [ebje]
‘tv’ [idje]
‘spoon’ [akje]
‘bread’ [ufje]
‘butter’ [ipje]
‘hand’ [ekji]
‘head’ [asji]
‘leg’ [utji]
‘heart’ [akji]
‘milk’ [edji]

Token training
(all words presented twice):

‘sofa’ [apu]
‘heart’ [akji]
‘house’ [ata]
‘bread’ [ufje]
‘table’ [eda]
‘milk’ [edji]
‘lamp’ [asa]
‘pot’ [ufu]
‘spoon’ [akje]
‘butter’ [ipje]