Optimality Theory and phonological acquisition

Paul Boersma, University of Amsterdam
&
Claartje Levelt, Leiden University

The past ten years have shown an ever-increasing revival of interest in phonological acquisition. The introduction of Optimality Theory (OT; Prince & Smolensky 1993) has played a major role in this revival. Both "markedness" and "constraints on outputs" have often figured in accounts of phonological development, and these concepts are central in OT. The most fundamental change is probably that OT has reinstated *continuity*: developmental grammars and final adult grammars have the same representational units and organizational principles. In the work by Jakobson (1941) and Stampe (1969) the development from an initial grammar to a final grammar had been pictured as a continuous process. For Jakobson development consisted of a gradual unfolding of a system of contrasts. For Stampe the initial grammar consisted of a set of unordered innate rules, while the final state consisted of an ordered subset of these rules. Development, then, consisted of suppressing the rules inappropriate to the adult target grammar and ordering the target-appropriate rules. In subsequent work, most notably by Smith (1973) and Kiparsky & Menn (1977) the child's developmental grammar(s) no longer had any principled relation to the target adult grammar.[1] In both works development mainly consisted of getting rid of a set of child-specific rules or strategies that filtered the target adult forms. Once this filter had been removed the target adult form and the child's production were identical. However, the ability to produce a target-appropriate form cannot be equaled to having acquired the adult grammar. Development in this view is thus non-continuous: the initial grammar, the "filter", does not develop into a mature grammar, but disappears without leaving a trace. In Parameter theory (Dresher & Kaye 1990, Fikkert 1994) there *is* continuity between developmental grammars and mature grammars: in the initial state of the grammar parameters have a default setting, and development consists of setting the appropriate parameters to their marked values. However, the successes of Parameter theory have largely remained domain-specific, in the sense that they have been confined to syllable structure and stress. In OT the basic idea is that constraints are innate and universal and come with an initial ranking where all markedness constraints (segmental, syllabic, and prosodic well-formedness constraints) outrank all faithfulness constraints (constraints requiring similarity between underlying and surface representations). The learner needs to acquire the language-specific ranking of these constraints. By subsequent rerankings, the initial

---

[1] Smith (1973) did argue that the child's so-called *realization rules* and general rules of phonology were made from the same stuff. The realization rules themselves were not related to phonological rules in adult grammars, however.

grammar gradually develops into the target-appropriate final grammar. The relation between child language phonology and phonology general has thus been reinstated – "kids avoid what adults avoid, only more so" (Smolensky, p.c. 1995) and this probably underlies the growing interest in phonological acquisition. This is not to say that all accounts of phonological acquisition are now OT accounts; the hypothesis space that OT offers is even regularly used to propose accounts of phonological development that lie outside of the theory proper. The field is moving in various directions, and interesting combinations of computational and empiricist, theoretical and experimental research are developing.

With this state-of-the-art paper in mind, a survey was sent out to a large group of people involved in research on phonological acquisition. They were asked, among other things, what *they* thought the "state of the art" of the field was, and in which directions the field should proceed. The response rate was high, and the responses were very informative. Most people agreed on the current strong influence of Optimality Theory on acquisition research, and this paper will therefore focus on OT. We will start out, in §1, by presenting OT basics. In §2, we will discuss acquisition research in the light of two important hypotheses of OT. Then, in §3, the focus is on learnability issues in OT. In §4, we will try to answer the question how the field should proceed, taking into account the respondents' main concerns.

1. Optimality Theory

1.1. Fundamentals of Optimality Theory

In the original version, an OT grammar consists of a set of innate, universal, violable constraints on *output* (surface) representations. There are two large sets of interacting constraints: *markedness* constraints, which demand structural unmarkedness of output representations, and *faithfulness* constraints, which demand that the output representations are similar to the *input* (underlying) representations. These many requirements are often in conflict. For example, markedness constraints prefer output representations that are unmarked, but if the input representation contains a marked configuration, faithfulness constraints require this marked structure to show up in the output. These and other types of conflict are resolved by constraint ranking. A higher-ranked constraint takes precedence over a lower-ranked constraint in determining the ultimate shape of the output. Depending on the language, violations of one constraint are considered to be worse than violations of some other constraint. The ranking of constraints is thus language-specific. All constraints are present in all grammars, although many of the constraints will usually have only subtle effects, or even no effect, on output shapes because of their low rankings. Constraints are violable, but only minimally so. An output candidate can violate a lower-ranked constraint in order to satisfy a higher-ranked constraint. The actual output results from a process whereby a set of output candidates – possible linguistic analyses of the input – is evaluated against the constraint hierarchy. The winning candidate, the *optimal output*, is the one which least violates the constraint hierarchy.

OT thus differs from previous derivational accounts in several important ways. Apart from the fact that there are no rules, and no intermediate representations, there are also no restrictions on underlying representations. This is called *Richness of the Base* (see also §3.3). It implies that language-specific restrictions on surface forms and segment inventories are never consequences of restrictions in the lexicon, but always have grammatical causes, i.e., they are the result of interactions of markedness and faithfulness constraints.

As mentioned in the introduction, constraints refer to segmental, syllabic and prosodic properties. In addition they can refer to the alignment of segments, syllables and higher-order prosodic units, and to the alignment of phonological and morphological units. Since these constraints are all connected through the ranking, interactions between for example segmental and prosodic structure can be accounted for in a non-stipulative way.

1.2. Acquisition as reranking of constraints

Optimality Theory makes two particular strong points for acquisition research. One point, already mentioned above, is the fact that OT has incorporated notions like "markedness" and "output constraint" which already played an important role in studies on child language phonology. Many aspects of child language phonology that researchers have always tried to, or wanted to, express *can* suddenly be expressed within the larger context of a formal theory. This further explains the renewed interest in the field: there is a fit between the empirical data and the theoretical vocabulary. The other particular strength of OT for acquisition research lies in the way it stakes out the hypothesis space. It is possible to formulate precise hypotheses and to test them. It turns out that based on this type of research some of the same conclusions that child language phonologists posited decades ago need to be drawn again. However, these conclusions now have a theoretical basis.

Since constraints are assumed to be innate and universal, acquiring an OT grammar should basically mean acquiring the language-specific ranking of these constraints. There are two questions in this respect: what is the initial state of an OT grammar, and how does it subsequently develop?

As for the issue of the initial state, early research in formal learnability (Tesar & Smolensky 1993 et seq.) had simply assumed that all constraints started out being unranked. Although the resulting learning algorithms were capable of acquiring a grammar, they were not designed to model the productions that the learner utters during the developmental period. In work based on empirical studies (e.g. Gnanadesikan 1995/to appear, Levelt 1995), it was pointed out that outputs are initially governed by markedness constraints, rather than by faithfulness constraints. This led to the proposal that in the initial state of the grammar, all markedness constraints outrank all faithfulness constraints, or "M >> F" for short. For formal learning algorithms this turned out to be a better starting point as well, since it partially circumvented the subset problem (see §3.3): by starting out with the most constricted system, a learner can reach the superset grammar by taking into account positive evidence only. Both sources of scientific evidence (namely empirical data

and learning algorithms) thus converged on an initial state M >> F (but see Hale & Reiss 1998 for an opposite proposal).

As for the subsequent reranking of the constraints, the general view became that learning was error-driven (see §3.1 and §3.2): when the learner detects a disparity between the output of her grammar and the adult target, she will lower the rankings of some or all of the constraints that favor her own output, and she may also raise the rankings of some or all of the constraints that favor the adult target. In early M >> F stages, this means that if the adult target form contains marked structure that is disallowed by a high-ranked markedness constraint in the learner's grammar, this markedness constraint tends to be demoted, and the corresponding faithfulness constraint may be promoted; this will sooner or later lead to some F >> M rankings. Unlike the default setting of a parameter, which has no effect at all once the parameter has been set to its marked value, a constraint is not rendered powerless once it has become low-ranked. Since the constraint remains in the grammar, its presence may still be felt under specific circumstances.

1.3. Underlying representations

What about the underlying representation, the input to the grammar? How does the learner establish the input to the OT grammar? This has been an issue in all theories: input-output pairings are needed to establish a grammar, while a grammar is needed to establish the input representation. Smolensky (1996a) argues that the learner will select the input representation that matches the adult output representation as the optimal input, even when the learner's grammar is still unlike the adult grammar. This follows from the combination of two OT tools, *Richness of the Base* and *Lexicon Optimization*. The example, after Smolensky, is the word *cat* /kæt/. The learner produces [kæ] for *cat*, because the markedness constraint NoCODA ("a syllable should not have a coda") outranks the faithfulness constraint MAX(seg) ("don't delete a segment") in his or her grammar. Tableau (1) shows how the child makes the crucial choice between [kæ] and [kæt] and at the same time rules out two other candidates with the faithfulness constraints DEP(seg) "don't insert a segment" and IDENT ("don't replace a segment").

(1)   *The child's production of* cat

/kæt/	NoCODA	DEP(seg)	IDENT	MAX(seg)
[kæt]	*!			
☞ [kæ]				*
[kæti]		*!		
[kæi]			*!	

In (1) the presumed input is /kæt/. The question is, did the learner indeed store this input representation when she heard an adult produce the word *cat* [kæt]? Assuming that the learner has indeed perceived [kæt], the argument proceeds as follows. The

perceived adult [kæt] is now evaluated by the same grammar used for production, this time to find the optimal input matching the output. See tableau (2).

(2)   *The child's comprehension of* cat

input candidates	perceived output	NOCODA	DEP(seg)	IDENT	MAX(seg)
☞ /kæt/	[kæt]	*			
/kæ/	[kæt]	*	*!		
/æ/	[kæt]	*	*!*		
/dɔg/	[kæt]	*		*!**	
/skæti/	[kæt]	*			*!*

Since [kæt] contains a coda consonant, the constraint NOCODA is always violated by this output, no matter what input is paired with it. This means that for comprehension, unlike for production, the NOCODA constraint is not decisive. The next constraint, the faithfulness constraint MAX, does prefer a single input: /kæt/, which is the only input-output pair that does not involve a MAX violation. From all the possible input candidates – *Richness of the Base* – the optimal one turns out to be the candidate that, all else being equal, is most faithful to the perceived output – this is called *Lexicon Optimization* (Prince & Smolensky 1993). As soon as the learner becomes aware of morpho-phonological alternations, certain input representations might need to be revised. While the input candidate /fiɔnt/ is for a long time the optimal input candidate for the perceived output [fiɔnt] '*dog*' in Dutch, it needs to be revised as soon as the learner relates the perceived output [fiɔndə] '*dogs*', to the perceived output [fiɔnt] '*dog*'.

In models of phonological acquisition it is often assumed that the child's perception is accurate. Many studies on infant speech perception (e.g. Eimas, Siqueland, Jusczyk & Vigorito 1971; Kuhl 1991; for an overview, see Jusczyk 1997) have indeed shown that infants can discriminate precisely their language-specific sound contrasts long before they start producing speech. These findings appear to support the assumption in most empirical studies on phonological acquisition in OT that the child's lexical representation is largely equivalent to an adultlike surface representation. However, other studies have shown that lexical representations are not always accurate. Hallé & Boysson-Bardies (1996), for instance, conclude that 11-months-old children have underspecified representations of familiar words. Recent studies by Stager & Werker (1997) and Werker & Stager (2000) find that sounds that could be distinguished by children around 14 months old in a pure speech perception task could not be distinguished in a word learning task. Since the onset of meaningful speech production is also somewhere between 11 and 14 months, it is not unlikely that in the early stages, lexical representations are un(der)specified in certain respects, and therefore less target-appropriate than assumed in much OT work. We will come back to this issue below.

Now that the basics of OT and the basic OT assumptions for acquisition have been discussed, we can turn to acquisition research proper.

2. Optimality Theory and acquisition research

By now, OT accounts have been presented for virtually every aspect of child phonology: sound substitutions and segmental development (Adam 2003; Dinnsen & McGarrity 1999; Dinnsen & O'Connor 2001; Dinnsen, O'Connor & Gierut 2001; Gilbers 2001; Gilbers & Van der Linde 1999; Joppen & Grijzenhout 2000; Inkelas & Rose 2003), prosodic structure, truncations, stress (Adam 2003; Demuth 1995a, 1995b, 1996; Lleó & Demuth 1999; Pater 1997; Ota 1999), cluster reduction (see references in §2.2.1), syllable structure (Joppen & Grijzenhout 1999; Levelt, Schiller & Levelt 2000; Levelt & Van de Vijver, to appear), consonant harmony (see references in §2.3), variation (Adam 2003; Dinnsen & McGarrity 1999; Gierut, Morrisette & Champion 1999; Pater & Werle 2001), the interaction between phonology and morphology (Adam 2003; Lléo 2001; Lléo & Demuth 1999), everything (Bernhardt & Stemberger 1998). However, rather than presenting a complete overview of OT acquisition research, we choose to focus on two important OT-based statements concerning acquisition. These are "Development = reranking," discussed in §2.1, and "Developmental grammars correspond to adult final grammars," discussed in §2.2. These statements will be illustrated with OT acquisition research. In §2.3 we will review work on a process that has been considered a typical child language process: consonant harmony.

2.1. Development = reranking

The basic premise is that markedness constraints outrank faithfulness constraints in the initial state of the grammar, and that development consists of reranking constraints. Reranking will stop when the outputs of the developing grammar and those of the target adult grammar are identical. A couple of studies have shown that developmental stages can indeed be captured by positing successive grammars in which the relative rankings of markedness and faithfulness constraints change from M >> F to, eventually, F >> M. We will illustrate development as reranking with work by Levelt, Schiller & Levelt (2000), Ota (1999) and Pater (1997).

2.1.1. Acquisition of syllable types

The core syllable consists of an onset consonant (C) followed by a vowel (V), i.e. CV. This is assumed to be the most unmarked syllable type, and in OT every deviation from this form therefore constitutes a violation of one of the syllabic markedness constraints. These constraints are:

(3) *Syllabic markedness constraints*

ONSET: a syllable should have an onset.
NOCODA: a syllable should not have a coda.
*COMPLEX-ONSET: a syllable should not have a complex onset.
*COMPLEX-CODA: a syllable should not have a complex coda.

In the initial state of the grammar these constraints all outrank faithfulness. In this stage, then, the only possible optimal outputs are CV outputs. This has indeed been

noticed time and again in early child language data. Levelt, Schiller & Levelt (2000) report on a study on the development of syllable types in longitudinal, naturalistic data of 12 children acquiring Dutch. The children that produced only one syllable type indeed produced CV syllables exclusively. This apparently confirms the M >> F point of departure for development. For Dutch, the final ranking is the reverse for syllable structure, F >> M, since every markedness constraint in (3) can be violated in the adult language. The language allows vowel-initial syllables (which violate ONSET), closed syllables (which violate NOCODA), and syllables with onset clusters or coda clusters (which violate the *COMPLEX constraints).

In the study by Levelt et al. it turned out that the different syllable types were acquired step by step, and that the order of acquisition was surprisingly similar across the 12 children. The only variation concerned the acquisition of consonant clusters. One group of children acquired complex onsets before complex codas, while another group acquired complex codas before complex onsets.

(4) *Order of acquisition of syllable types*

$$\begin{array}{ccc} & \text{group A: CVCC > VCC > CCV > CCVC} & \\ \text{CV > CVC > V > VC} & & > \text{CCVCC} \\ & \text{group B: CCV > CCVC > CVCC > VCC} & \end{array}$$

The general direction of development can indeed be captured by intermediate grammars that link the initial to the final grammar by demoting the markedness constraints below faithfulness one by one:

(5) *Developmental grammars*

   Stage 1: { *COMPLEX-O, *COMPLEX-C, ONSET, NOCODA } >> **FAITH**
     ⟹ allows CV outputs only.
   Stage 2: { *COMPLEX-O, *COMPLEX-C, ONSET } >> **FAITH** >> NOCODA
     ⟹ allows CV, CVC.
   Stage 3: { *COMPLEX-O, *COMPLEX-C } >> **FAITH** >> { ONSET, NOCODA }
     ⟹ allows CV, CVC, V, VC.
   Stage 4a: *COMPLEX-O >> **FAITH** >> { *COMPLEX-C, ONSET, NOCODA }
     ⟹ allows CV, CVC, V, VC, CVCC, VCC.
   Stage 4b: *COMPLEX-C >> **FAITH** >> { *COMPLEX-O, ONSET, NOCODA }
     ⟹ allows CV, CVC, V, VC, CCV, CCVC.
   Stage 5: **FAITH** >> { *COMPLEX-O ,*COMPLEX-C, ONSET, NOCODA }
     ⟹ allows all syllable types.

There are some more subtle developments that the above developmental grammars fail to capture. For example, it was found that for some children the syllable type VC occurred in the data at a later stage than the syllable type V. This is not expected: as soon as both NOCODA and ONSET are dominated by FAITH both V and VC should be possible. It is not clear if indeed every development in the data should be taken to reflect a grammatical development. The findings could turn out to be artifacts of the data collection. However, these more subtle developments can be accounted for in the grammar if *local conjunction* of constraints is invoked. The syllable type VC is more

marked than either V or CVC in the sense that it violates both ONSET and NOCODA. In the evaluation of a VC candidate for a VC input, in a grammar where FAITH outranks NOCODA and ONSET, these two markedness violations do not play a role, since VC is faithful. In OT there is the possibility to combine the constraints ONSET and NOCODA in a new conjoined constraint, which is ranked above the individual constraints (Smolensky 1993). The new, conjoined constraint ONSET&NOCODA is violated when both parts of the conjunction are violated. In (6) it is illustrated how a grammar with ONSET&NOCODA rules out VC while allowing for V and CVC:

(6) *Conjoined constraints in the grammar*

/VC/	ONSET&NOCODA	FAITH	ONSET	NOCODA
VC	*!		*	*
V		*	*!	
☞ CVC		*		*

/V/	ONSET&NOCODA	FAITH	ONSET	NOCODA
VC	*!	*	*	*
☞ V			*	
CVC		*!*		*

/CVC/	ONSET&NOCODA	FAITH	ONSET	NOCODA
VC	*!	*	*	*
V		*!*	*	
☞ CVC				*

In OT this move has consequences, however. Since constraints are supposed to be universal, what is the status of the conjoined constraint in this respect? If it is a universal constraint, we expect to find more evidence for the presence of this constraint in grammars. If no such evidence can be found, the constraint might reflect a child-specific, transient processing problem. The question is then if such problems should be accounted for in the phonological grammar. However, with or without conjoined constraints, development can be accounted for in terms of reranking, whereby markedness constraints are demoted below a competing faithfulness constraint one by one.

2.1.2. Acquisition of prosodic structure

Our next example of development as reranking comes from the dissertation of Ota (1999) and concerns the development of word-internal prosodic structure. Children often produce truncated forms of adult targets. In most accounts this phenomenon has been related to the fact that the child maps the adult target onto some specific prosodic shape, like the minimal word. In OT the prosodic shape of a word is

regulated by a set of markedness constraints. Ota accounts for the initial prosodic shape and subsequent developments of child forms in Japanese using the following set of markedness constraints:

(7) *Prosodic markedness constraints*

FTBIN: feet are binary at some level of analysis (syllable, mora).
ALIGNFT-L: every foot must be left-aligned within a prosodic word.
PARSE-σ: every syllable must be parsed (contained) in a foot.
TROCHAIC: every accented syllable must be left-aligned within a foot.

These constraints compete with the faithfulness constraints:

(8) *Faithfulness constraints*

MAX(seg): every segment in the input must have a correspondent in the output.
DEP(μ): every mora in the output must have a correspondent in the input.
IDENT(σ́): every output correspondent of an accented syllable in the input must be accented.

If the OT assumption of an initial M >> F ranking is correct, we expect that in an early stage all four markedness constraints are ranked at the top, so that optimal candidates will always have the shape of a binary trochaic foot. This is indeed what Ota finds in his Japanese child language data. Longer adult target forms are truncated, while monomoraic adult target forms are augmented in the children's productions. Children are always faithful, though, to the accented syllable, if there is one. Examples are in (9).

(9) *Augmentations and truncations in child Japanese*

	input	gloss	winning output	violation of
a.	/mé$_μ$/	'eye'	(me$_{μμ}$)	DEP(μ)
b.	/bánana/	'banana'	(bána)	MAX(seg)
c.	/mɯzɯkaçiː/	'difficult'	(káçi)	MAX(seg)
d.	/takái/	'high'	(kái)	MAX(seg)

In (10) we see the tableau for the form /takái/ (from Ota 1999). Parentheses mark foot boundaries.

(10) *Truncation in Japanese*

/takái/	TROCHAIC	PARSE-σ	IDENT(σ́)	MAX(seg)
(takái)	*!			
ta(kái)		*!		
(tákai)			*!	
☞ (kái)				**

9

In the adult state of the grammar, the markedness constraints are ranked below the faithfulness constraints. However, Ota finds evidence for an intermediate stage, where TROCHAIC has been demoted below the faithfulness constraint MAX, while the other markedness constraints still outrank this constraint. At this stage, the input form /takái/ is produced faithfully, while longer forms are still truncated. The developments are, then, as in (11). Ota does not present all the constraints in a single grammar, but only presents rankings for which ranking arguments are found in the data.

(11) *Rerankings in the grammar*

With respect to DEP:
Stage 1: FTBIN >> **DEP**
    ⇒ augmentation of monomoraic target forms.
Stage 2: **DEP** >> FTBIN
    ⇒ faithful monomoraic outputs are possible.

With respect to MAX:
Stage 1: { FTBIN, ALIGNFT-L, IDENT(ó), TROCHAIC } >> **MAX**
    ⇒ truncation of longer adult target forms to binary trochaic feet.
Stage 2: { FTBIN, ALIGNFT-L, IDENT(ó) } >> **MAX** >> TROCHAIC
    ⇒ truncation, faithful right-aligned accent possible.
Stage 3: **MAX** >> { FTBIN, ALIGNFT-L, IDENT(ó), TROCHAIC }

Development of prosodic structure can thus also be captured by a gradual reranking of constraints.

2.1.3. Development is reranking of constraints, not getting rid of constraints

Stampe (1969) proposed that children are born with a large set of rules. The language learner needs to suppress the rules that are not operative in the target language. In OT the idea is that children are born with a set of constraints, which need to be reranked. Unlike the suppressed rules, reranked constraints remain present in the grammar. Why don't constraints disappear from the grammar, like they used to do in pre-OT accounts of child language phonology? Minimal violation is the keyword here. A constraint is not violated unless not violating this constraint leads to the violation of a higher-ranked constraint. But even in this case, the optimal candidate will be the candidate that only minimally violates the constraint. An output candidate with more violations of the constraint will loose out to the candidate that minimally violates the constraint. The presence of this dominated constraint in the grammar is thus required.

Pater (1997) sets out to show that constraints that start out being undominated in the child's grammar still play a role in the mature grammar even though they have ended up being dominated by other constraints.

Pater discusses truncations in English child language, and it should not come as a surprise that these truncations look very similar to the truncations of Japanese children we just discussed. The initial grammar proposed for English is therefore already familiar: the combination of the high-ranked prosodic markedness constraints

ALIGNFT-L, PARSE-σ, and FTBIN, which dominate the faithfulness constraint MAX(σ) force the optimal output to equal a minimal word, i.e. a single trochaic foot.

(12) *Truncation grammar*

/hippopotamus/	ALIGNFT-L	PARSE-σ	FTBIN	MAX(σ)
(hippo)(pota)mus	*(!)*	*(!)		
(pota)mus		*!		**
(potamus)			*!	**
(hippo)(pomus)	*!*			*
☞ (pomus)				***

In adult English, MAX(σ) dominates at least ALIGNFT-L and PARSE-σ, as required by the faithful output for /hippopotamus/: winning output candidates can contain both more than one foot, and unfooted syllables. Is there any evidence for minimal violation of ALIGNFT-L and PARSE-σ in English, which would argue for their presence in the adult grammar? Pater finds two effects of these dominated constraints.

Firstly, although unfooted syllables are allowed in English, the optimal candidate is the candidate that has parsed as many syllables as possible into feet. This has the effect of secondary stress in the language, as in the word *àpalàchicóla*. As can be seen below in (13) every foot except the first one constitutes an ALIGNFT-L violation – a violation mark is given for every syllable that lies in between the foot and the left word edge – and this constraint must therefore be dominated by PARSE-σ in the adult grammar:

(13) *Minimal violation of PARSE-σ*

/apalachicola/	MAX(σ)	PARSE-σ	ALIGNFT-L
(cóla)	*!***		
(ápa)lachicola		*!***	
☞ (àpa)(làchi)(cóla)			******

The effect of ALIGNFT-L can be seen in words like *tàtamagóuchi, àbracadábra* and *wìnnepesáuki*, where a trisyllabic sequence precedes the main stress. The position of the foot carrying the secondary stress has to be determined by ALIGNFT-L. A *right*-alignment constraint like ALIGNFT-R would have preferred the *abràcadábra* candidate, but this constraint is clearly outranked by ALIGNFT-L.

(14) *Minimal violation of ALIGNLEFT*

/abracadabra/	MAX(σ)	PARSE-σ	ALIGNFT-L	ALIGNFT-R
(dábra)	*!**			
a(bràca)(dábra)		*	****!	**
☞ (àbra)ca(dábra)		*	***	***

Effects like these show that output constraints that were fully satisfied in the initial state of the grammar are not shut off during the acquisition process. They are part of the adult final grammar.

2.1.4. Conclusions about development as reranking

The above examples of development as reranking have shown that systematic permutations of a single, restricted set of constraints can account for empirical developmental data. The permutations are systematic in the sense that initially high ranked markedness constraints are demoted below competing faithfulness constraints. The constraints are not child specific, since more subtle effects of the same markedness constraints can be found in the adult target language. At least in these respects, then, phonological development is a continuous process. However, it could still be the case that the specific constraint rankings of developmental grammars are child-specific. In the next paragraph we will discuss work showing that developmental grammars are not unique to language learners, and have final state counterparts.

2.2. Developmental grammars correspond to adult final grammars

It follows from the basic principles of OT that developmental grammars are not principally different from adult grammars. If children avoid what adults avoid it should be possible to pair any developmental grammar with some final state adult grammar. Of course no adult final grammar will resemble a developmental grammar in its entirety, but there should be corresponding subgrammars. This prediction plays a role in most of the work on acquisition within the OT framework, and we will illustrate this with analyses of Gnanadesikan (1995/to appear) and Levelt & Van de Vijver (to appear).

2.2.1. Cluster reduction in child English and Sanskrit

Gnanadesikan (1995/to appear) is among the first who accounted for child language data within OT, systematically comparing developmental states of the grammar of a child acquiring English to final grammars of Sanskrit, Kolami and Navajo. Here we illustrate the relation between Sanskrit and child English.

Both Sanskrit and developmental English at a certain stage avoid onset clusters. This is captured by a high-ranked markedness constraint *COMPLEX. An additional similarity between the two languages is that the input representation does

contain complex onsets. Sanskrit does allow onset clusters, but avoids them in reduplication: *prach*, *sna:*, *sru*, etc. become *pa-prach*, *sa-sna:* and *su-sru* in the perfect. The adult outputs of English perceived by the child contain onset clusters, and the optimal inputs will thus contain them too. This means that both in Sanskrit reduplication and in child English the grammar enforces an unfaithful output. The faithfulness constraints that are violated are slightly different. In Sanskrit the reduplicant is unfaithful to its base form. This is a form of Output-Output (O-O) unfaithfulness. O-O faithfulness constraints enforce corresponding forms in morphological paradigms. In child English it is Input-Output (I-O) faithfulness that is compromised. However, the grammars show a similar markedness effect in the solution to the COMPLEX problem.[2] Both languages opt for cluster reduction. Which segment of the cluster is retained in the output is determined by markedness constraints that can be violated otherwise, another example of minimal violation. There is a set of pre-ranked markedness constraints that together constitute the sonority hierarchy. To present a simple version here, the following fixed ranking of constraints aims at minimizing the sonority of onsets:

> *ONS/VOWEL >> *ONS/GLIDE >> *ONS/LIQUID >>
> >> *ONS/NASAL >> *ONS/FRICATIVE >> *ONS/PLOSIVE

A plosive onset thus violates the constraint *ONS/PLOSIVE, but since this constraint will usually be quite low-ranked in the grammar, its effect will hardly ever be felt. In both the Sanskrit grammar and the child English grammar all types of onsets, except for vowels, are allowed, and are thus outranked by faithfulness constraints. However, their presence is suddenly felt when one of the cluster segments needs to be deleted: the onset segment which violates the highest ranked markedness constraint, i.e. the most sonorous segment, will be will be deleted, even though this element is otherwise allowed. This phenomenon is called the *Emergence of the Unmarked*, and is one of the hallmarks of OT. Sanskrit and the specific type of child English discussed by Gnanadesikan at some point thus have a similar constraint ranking: *COMPLEX >> FAITH (I-O or O-O) >> *ONS/GLIDE >> *ONS/LIQUID >> *ONS/NASAL >> *ONS/FRICATIVE >> *ONS/PLOSIVE.

(15) *Similar grammars for Sanskrit and child English*

(a) *Sanskrit*

/sna:+RED/	I-O FAITH	*COMPLEX	O-O FAITH	*ONS/ GLIDE	*ONS/ LIQUID	*ONS/ NAS	*ONS/ FRIC	*ONS/ PLOS
sna:-sna:		**!				**	**	
na:-sna:		*	*			**!	*	
☞ sa:-sna:		*	*			*	**	
sa:-sa:	*!						**	

(b) *Child English*

/sno:/	*COMPLEX	I-O FAITH	*ONS/ GLIDE	*ONS/ LIQUID	*ONS/ NAS	*ONS/ FRIC	*ONS/ PLOS
sno:	*!				*	*	
no:		*			*!		
☞ so:		*				*	

In dealing with *COMPLEX, the general pattern in development is to reduce the cluster to the least sonorant segment. However, in child language other patterns occur as well, due to interactions with other markedness and faithfulness constraints (e.g., the candidate [no:] can win because of contiguity constraints or because of /s/ extrasyllabicity). Several in-depth studies on the development of clusters within an OT framework have been carried out. We would like to refer the interested reader to the work by Barlow (1997), Barlow & Pater (to appear), Jongstra (2003), and Kirk & Demuth (2003).

2.2.2. Syllable types in developmental- and cross-linguistic grammars

The research reported on in Levelt & Van de Vijver (to appear) is specifically set out to test the hypothesis that there is a 1:1 relationship between developmental grammars and adult final state grammars. They compare the developmental grammars linking G_i to the final state grammar for syllable types in Dutch from the Levelt et al. study to final state grammars for syllable type. This cross-linguistic information comes from a survey of syllable types in the world's languages by Blevins (1995). As a first finding, it turns out that the syllable type structure of every language discussed in Blevins can be captured by a specific ranking of the syllabic markedness constraints presented in (1) above vis-à-vis a faithfulness constraint FAITH. Moreover, every permutation of the constraints – markedness vis-à-vis faithfulness - corresponds to an empirically attested grammar. The set of possible grammars that results from all the potential rankings of a set of constraints is called a *factorial typology*. This is illustrated in (16):

(16) *Factorial typology of syllabic constraints*

Grammar	*Language*
{ ONSET, NOCODA, *COMPLEX-O, *COMPLEX-C } >> **FAITH**	Hua
{ NOCODA, *COMPLEX-O, *COMPLEX-C } >> **FAITH** >> ONSET	Cayuvava
{ ONSET, *COMPLEX-O, *COMPLEX-C } >> **FAITH** >> NOCODA	Thargari
{ ONSET, NOCODA, *COMPLEX-C } >> **FAITH** >> *COMPLEX-O	Arabela
{ ONSET, *COMPLEX-C } >> **FAITH** >> { NOCODA, *COMPLEX-O }	Sedang
{ NOCODA, *COMPLEX-C } >> **FAITH** >> { ONSET, *COMPLEX-O }	Mazateco
{ *COMPLEX-O, *COMPLEX-C } >> **FAITH** >> { ONSET, NOCODA }	Mokilese
{ ONSET, *COMPLEX-O } >> **FAITH** >> { NOCODA, *COMPLEX-C }	Klamath
ONSET >> **FAITH** >> { NOCODA, *COMPLEX-C, *COMPLEX-O }	Totonac
*COMPLEX-C >> **FAITH** >> { ONSET, NOCODA, *COMPLEX-O }	Spanish
*COMPLEX-O >> **FAITH** >> { ONSET, NOCODA, *COMPLEX-C }	Finnish
FAITH >> { ONSET, NOCODA, *COMPLEX-O, *COMPLEX-C }	Dutch

14

Repeating here the developmental grammars for syllable structure that were presented in (5) above, it can be confirmed that indeed all of these developmental grammars correspond to some attested final grammar:

(17) *Developmental grammars for syllable type*

Stage 1: { *COMPLEX-O, *COMPLEX-C, ONSET, NOCODA } >> **FAITH**
Stage 2: { *COMPLEX-O, *COMPLEX-C, ONSET } >> **FAITH** >> NOCODA
Stage 3: { *COMPLEX-O, *COMPLEX-C } >> **FAITH** >> { ONSET, NOCODA }
Stage 4a: *COMPLEX-O >> **FAITH** >> { *COMPLEX-C, ONSET, NOCODA }
Stage 4b: *COMPLEX-C >> **FAITH** >> { *COMPLEX-O, ONSET, NOCODA }

The initial state grammar corresponds to the grammar for Hua, the grammar of stage 2 corresponds to the grammar for Thargari, the grammar of stage 3 to the grammar for Mokilese, the grammar of stage 4a to the grammar for Spanish, and the grammar of stage 4b corresponds to the grammar for Finnish. This confirms the OT hypothesis that child grammars and adult grammars are made of the same stuff.

An interesting question is why there is hardly any variation in the learning paths taken by the Dutch children. For example, for all children the grammar at stage 2 corresponds to the grammar of Thargari. What prevents the choice for Cayuvava or Arabela? These possibilities are neither excluded by the grammar nor by the theory. Levelt & Van de Vijver conclude that the frequency of syllable types in the target language determines the choice for one developmental step over another. In Dutch, CVC, the syllable type that would become possible when NOCODA would be demoted, is highly frequent. It occurs over 20 times more often than CCV, which would become an available output if *COMPLEX-O were demoted (the Arabela option) and over 8 times more often than V, which would become available if ONSET were demoted (the Cayuvava option). The learner thus makes an economical decision. In the one case where variation does occur (one group demotes *COMPLEX-O first, the Spanish option, while the other group demotes *COMPLEX-C first, the Finnish option), frequency is not decisive. The resulting possibilities created by the two new grammars occur with almost equal frequencies in the language. The hypothesis that follows from this account is that we expect to find different learning paths for target languages with similar grammars but with different distributions of syllable types.

2.2.3. Conclusions about developmental grammars and final grammars

The work we just presented shows that it is possible to assume an even higher degree of continuity between child- and adult grammars than could be concluded from §2.1. Not only can child grammars be described using the same primitives and principles used for adult grammars, they have also been shown to actually correspond to sub-grammars of attested adult languages.

After this jubilant conclusion, we will now turn to a topic in child language phonology that has always challenged continuity, *consonant harmony*.

2.3. Consonant harmony

In *consonant harmony* (CH), two non-adjacent consonants in a word share a phonological feature. In the most salient type of CH in child language, the consonants share a primary place of articulation (PoA) feature, Labial, Coronal or Dorsal. Well-known examples are in (18):

(18) *Examples of consonant harmony*

duck /dʌk/ → [gʌk]
dog /dɔg/ → [gɔg]
zeep 'soap' /zeːp/ → [feːp]
neef 'cousin' /neːf/ → [meːf]

This phenomenon challenges continuity because it is not attested in this specific form in adult languages. In adult languages non-adjacent consonants can share secondary PoA features, but never primary PoA features (Hansson 2000). Given OT's strong inclination towards continuity, it is interesting to discuss OT accounts of CH from this viewpoint. We will concentrate on the accounts of Goad (1997), Pater (1997), Pater & Werle (2001), Rose (2001) and Fikkert & Levelt (2002).

2.3.1. Constraints that could explain consonant harmony

Given the fact that CH is present in child language, which is characterized by its general unmarked structure, the consensus is that CH output forms must be triggered by a high-ranked markedness constraint in the grammar. In the OT accounts, two different approaches to CH can be found. In one, CH results from a high-ranked markedness constraint that requires PoA agreement between consonants in a word. In Pater (1997) we find the constraint REPEAT, and Pater & Werle (2000) posit the constraint AGREE. In the other approach CH results from a constraint that requires a relation between a specific PoA feature and a specific edge or head of some domain: Goad (1997) and Fikkert & Levelt (2002) posit alignment constraints, Rose (2000) a licensing constraint. In this approach consonant harmony is an artifact of alignment or licensing.

The CH constraint interacts with constraints that require a faithful rendition of the input PoA specification in the output, and, in the analyses of Goad and Pater&Werle, with a constraint against a representational gap, NOGAP. This constraint enforces locality: a spreading feature cannot skip a potential docking site such as the intervening vowel. Below in (19) and (20) are two examples of CH-triggering grammars, one from Pater & Werle and one from Goad. In Pater & Werle's account the interacting constraints are:

AGREE-L-[DORS]: a consonant preceding a dorsal must be homorganic with it.
NOGAP: feature-linked segments must be adjacent at the root node level.
FAITHLAB, FAITHCOR, FAITHDORS: input and output feature specifications have to correspond.

(19) *If the consonant harmony constraint is* AGREE

/tek/ 'take'	FAITHDORS	AGREE-L	NOGAP	FAITHLAB	FAITHCOR
tek		*!			*
tet |/ Cor	*!				
☞ kek \ / Dors			*		*

Goad argues for a copy analysis, instead of a spreading analysis. In the winning output candidate a feature has been inserted, replacing the original feature. This of course induces violations of faithfulness constraints. Her analysis requires the following constraints (slightly modernized here):

ALIGN(DORS/LAB/COR): align the PoA feature with the left edge of the foot.
NOGAP
MAX(LAB/COR/DORS): no deletion of an input specification Lab, Cor or Dors.
DEPFEAT: no insertion of a PoA feature in the output.

(20) *If the consonant harmony constraint is* ALIGN

/tek/ 'take'	NOGAP	MAX DORS	MAX LAB	ALIGN DORS	ALIGN LAB	MAX COR	ALIGN COR	DEP FEAT
tek				*!				
tet / \ Cor Cor		*!						*
kek \ / Dors	*!					*		
☞ kek / \ Dors Dors						*		*

We will not go into the details of the proposed analyses here, but focus on the continuity problem instead. Why are the grammars in (19) and (20) never attested in adult languages? This is puzzling, especially in the light of the findings we presented in §2.1 and §2.2. We might need to conclude, then, that certain aspects of child phonology are child-specific after all. In OT accounts we find two ways in which child-specific aspects have been incorporated in the grammar: in the contents of an otherwise universal constraint and in child-specific constraints. Child-specific input representations will be discussed in §4.

2.3.2. Universal constraint templates with child-specific content

Rose (2000) presents an analysis of CH in terms of constraints he argues are operative in adult languages too. In this case, the problem for continuity would amount to the apparent child-specific ranking of the constraints. However, on closer inspection of the CH triggering constraints, it turns out that while the constraint scheme might indeed be universal, the specific content of this scheme is child-specific. This entails that the character of a constraint can change during development.

A proposal along these lines is found in Pater & Werle (2001). Pater and Werle argue that their CH-triggering constraint, AGREE, is operative in adult languages too, but on a different domain. In child language the domain is the word, while in adult language the domain is reduced to string-adjacent consonants. The adult AGREE constraint triggers well-attested cases of place assimilation in adjacent consonants. Stampe (1969) too, allowed for rule modification next to rule suppression in development. In this approach it appears that constraints are still innate. The interesting question is then how and when the specific contents of a constraint are determined.

2.3.3. Child-specific constraints

Pater (1997), Goad (1997) and Fikkert & Levelt (2002) all admit child-specific constraints in the grammar. They are all aware of the consequences of this move, however. In Pater's words, an account has to be given for constraint genesis and for constraint extinction.

According to Goad the motivation for the presence of the CH triggering constraint is functional: it reduces the number of articulatory instructions required to produce the word. In our discussion of locally conjoined constraints we already asked whether such processing problems should be dealt with in the phonological grammar (see Hale & Reiss). An argument in favor of this position is that the CH constraint has been shown to interact with other constraints in the grammar.

Fikkert & Levelt (2002) present an account of the emergence of a CH triggering constraint in the developing grammars of children acquiring Dutch. They first observe that CH forms are initially not present in the child language data. On the contrary, children are initially pretty faithful to the PoA structure of the adult target forms they produce. However, their initial lexicon is very restricted, and is built up in a systematic way. It appears that as soon as children are able to analyze words into segments, they select vocabulary items that have a highly frequent pattern of PoA features. In Dutch these are words of the form Labial consonant-vowel-Coronal consonant, abbreviated PvT. In a short time the child's lexicon contains quite a number of these PvT words, which are faithfully produced. PvK words come in next, and again these words are produced faithfully. Then, unfaithful CH forms appear in the data, for TvP and KvP adult targets. What has happened? According to Fikkert & Levelt, the child has analyzed her own lexicon, which contains all these PvT and PvK forms (and no TvP and KvP forms). The child's conclusion is that Labial is situated at the left edge of the word. This conclusion becomes phonologized: a constraint requiring Labial to be licensed by the left edge, [LAB, emerges in the grammar in a high-ranked position. From this moment on, the grammar favors unfaithful CH

candidates over faithful TvP and KvP candidates that have not properly licensed their feature Labial.

To conclude, it might be necessary to take a somewhat less restrictive OT approach to language acquisition. However, this does not need to affect the basic idea about continuity: children still also avoid what adults avoid. Grammars are similar, and there is a continuous development from an initial grammar to a final grammar. But in addition to that there can be transient constraints in developmental grammars. The child's own lexicon could be a source for emerging constraints, and in this sense, these constraints are lexicon-specific, rather than child-specific. It can be hypothesized that as soon as the lexicon has expanded beyond a certain magnitude, it becomes impossible to generalize over the lexicon.

3. Learnability

3.1. Learning from fully known input-output pairs

In nearly all work in Optimality Theory outside acquisition, the typical tableau contains a single underlying form as its input, and multiple surface forms as its output candidates. Such tableaus can be seen as *mapping* an underlying form to a surface form, and can therefore be regarded as modelling the *production* process. This view of the grammar thus consists of a single process (production) that relates two representations (underlying and surface), as summarized in (21).

(21) *The production-only model of phonology*
production: [*underlying form*]$_{UF}$ → [*surface form*]

If all underlying forms and all corresponding surface forms of a language are known, it is possible to derive a constraint ranking for production. For instance, if an underlying form is [σ σ σ]$_{UF}$ (three syllables in a language without underlying stress specifications), and the corresponding surface form is [(σ σ́) σ] (an iambic foot followed by an extrametric syllable), then we can derive that this language is iambic, i.e. that the constraint IAMBIC outranks the constraint TROCHAIC. Whether feet are left-aligned or right-aligned in the phrase cannot be derived from the surface form [(σ σ́) σ] alone: the language could be left-aligning, or it could be right-aligning with final-syllable extrametricality. But as soon as we see that this language also has the underlying-surface pair [σ σ σ σ]$_{UF}$ – [(σ σ́) σ σ], we can conclude that this language has left-aligned feet, i.e. that the constraint ALIGNFT-L outranks the constraint ALIGNFT-R.

Several general algorithms have been devised for finding an appropriate constraint ranking when underlying and surface forms are given. The first such algorithm was *Batch Constraint Demotion* (BCD; Tesar & Smolensky 1993; Tesar 1995). By using knowledge of all the underlying–surface pairs and knowledge about what surface forms violate what constraints, this algorithm, when applied to our example, will correctly establish the ranking of TROCHAIC below IAMBIC, and of ALIGNFT-R below ALIGNFT-L. Although this algorithm had no claim of being a good

model of the acquisition of constraint rankings by language-learning children, it was natural to ask how it could be modified so as to become an acquisition model. The first step in that direction was *Error-Driven Constraint Demotion* (EDCD; Tesar 1995; Tesar & Smolensky 1998, 2000), in which learning proceeds *on line*, i.e. by processing single underlying-surface pairs as they are presented to the learner; the only 'memory' that this algorithm has, resides in the ranking of the constraints: unlike BCD, EDCD needs no memory of previous underlying-surface pairs, and may for that reason be a better model of how actual children behave. As an example, consider a child who tries to acquire the above left-aligning iambic language, but currently has a trochaic right-aligning grammar. If we ignore extrametricality and other constraints, the production tableau will look as in (22).

(22) *First input-output pair for EDCD*

[σ σ σ σ]$_{UF}$	TROCHAIC	IAMBIC	ALIGNFT-R	ALIGNFT-L
[(ó σ) σ σ]		*	*!*	
√ [(σ ó) σ σ]	*!		**	
[σ (ó σ) σ]		*	*!	*
[σ (σ ó) σ]	*!		*	*
☞ [σ σ (ó σ)]		*		**
[σ σ (σ ó)]	*!			**

The underlying form is [σ σ σ σ]$_{UF}$, i.e. a sequence of four syllables not marked for stress (let's assume that this language has no lexical stress). In this example, the highest ranked constraint is TROCHAIC, which rules out the two candidates with iambic feet. The choice between the two remaining candidates is determined by the gradient constraint ALIGNFT-R, which rules out the candidates in which the foot is not aligned with the right edge of the phrase. The output of the child's grammar will be [σ σ (ó σ)], i.e. a right-aligned trochaic foot preceded by two extrametrical syllables. This winning candidate is depicted by the pointing finger. Now suppose we tell the child that the actual form in the target language is [(σ ó) σ σ], i.e. a left-aligned iambic foot followed by two extrametrical syllables. In the child's tableau, we depict this form with a check mark (√). Now that the child's form and the adult form are different, EDCD will take action. It first looks for the highest ranked constraint that prefers the adult form to the child's form. This is IAMBIC. All the constraints that are ranked above it and prefer the child's form (in this case, only TROCHAIC) are then *demoted* to a ranking below this pivotal constraint. TROCHAIC thus ends up being ranked equally high as ALIGNFT-R. This new grammar can subsequently be confronted with new language data, for instance with the trisyllabic adult underlying form [σ σ σ]$_{UF}$ and the corresponding adult surface structure [(σ ó) σ]. This is shown in (23).

(23) *Second input-output pair for EDCD*

[σ σ σ]$_{UF}$	IAMBIC	TROCHAIC	ALIGNFT-R	ALIGNFT-L
[(ɔ́ σ) σ]	*!		*	
√ [(σ ɔ́) σ]		*	*!	
[σ (ɔ́ σ)]	*!			*
☞ [σ (σ ɔ́)]		*		*

In the comparison between the adult form and the child's form, the pivotal constraint is ALIGNFT-L, and ALIGNFT-R is the only constraint that has to be moved below it. This second EDCD step leads to an appropriate adultlike ranking, namely that of an iambic left-aligning language. The grammar now works correctly for all possible inputs [σ σ]$_{UF}$, [σ σ σ]$_{UF}$, [σ σ σ σ]$_{UF}$, [σ σ σ σ σ]$_{UF}$, and so on. So this example was a quick success, in which only two informative input-output pairs were needed (the attentive reader may have noted that chewing twice on the quadrisyllabic input-output pair would have led to successful acquisition with a single informative pair). Tesar and Smolensky show that if the learner is given enough randomly selected pairs of underlying form and surface structure, EDCD will eventually change any initial grammar in such a way that it assigns correct surface structures to all underlying forms. In other words, EDCD is guaranteed to succeed when full structural descriptions of surface forms are given.

Another step towards realism was taken by the *Gradual Learning Algorithm* (GLA; Boersma 1997, Boersma & Hayes 2001), in which the learner typically shows variation in the choice of her surface forms, and in which learning can be characterized by smooth learning curves. This algorithm differs from EDCD in the way the constraint are reranked in case of a mismatch between the child's form and the adult form. The algorithm assumes a grammar model in which all constraints are ranked along a continuous scale rather than being just ordered. The ranking of all the constraints that prefer the learner's form (in the case of tableau (2), TROCHAIC and ALIGNFT-R) is lowered by a small amount along this ranking scale, and the ranking of all the constraints that prefer the adult form (in our case, IAMBIC and ALIGNFT-L) is raised by the same amount. After the learner has received many input-output pairs, the rankings of IAMBIC and TROCHAIC will have reversed, as will the rankings of ALIGNFT-L and ALIGNFT-R.

For cases that do not involve optionality, both EDCD and GLA are *convergent*, i.e., they lead to a correct ranking of the constraints, if such a ranking exists for the given set of underlying-surface pairs (for EDCD, Tesar & Smolensky have rigorously proved this; for GLA, there is no rigorous proof, but no counterexample has been found). The existence of convergent learning algorithms is evidence for the appropriateness of Optimality Theory as a model of language typology, since the main alternative framework, namely parameter setting, does not come with a convergent algorithm (Gibson & Wexler 1994).

The GLA has been used for successful computer simulations of attested acquisition orders. Boersma & Levelt (2000) replicated the syllable type acquisition described by Levelt, Schiller & Levelt (2000), discussed above in §2.1.1, and Curtin

& Zuraw (2002) replicated a more complicated example of the acquisition of syllable structure in Dutch.

3.2. Learning from partially known input-output pairs

Neither EDCD nor GLA can be realistic models of language acquisition, because they require the learner to have access to fully structured adultlike surface forms, which are in reality generally hidden from the learner. This was stressed in a developing series of papers by Tesar (1997, 1998, 1999, 2000), which was ultimately completed as chapter 4 of Tesar & Smolensky (2000). In our example of an iambic left-aligning language environment, the real child only hears unstructured *overt forms* such as the syllable sequences [σ σ́ σ]$_{OF}$ and [σ σ́ σ σ]$_{OF}$, not the corresponding full phonological structures [(σ σ́) σ] and [(σ σ́) σ σ], i.e., she may be able to hear which syllable is stressed but the foot structure is hidden; and underlying forms are not in general trivially inferrable from the overt form either. This means that the language-learning child must somehow create both the underlying forms and the surface structures from the overt forms. If the language environment is a left-aligning iambic language, the child must ultimately learn to map the overt form [σ σ́ σ]$_{OF}$ to the surface structure [(σ σ́) σ], whereas if the child grows up in a right-aligning trochaic language, she must learn to map it to the surface structure [σ (σ́ σ)]. Therefore, the mapping from overt form to surface structure is non-trivial and language-dependent, and has to be modelled explicitly by any acquisition model that makes claims to realism. A grammar model that takes acquisition seriously thus has to distinguish at least three types of forms (underlying, surface, and overt), and two processes: production and comprehension. We summarize this in (24).

(24) *The bidirectional model of phonology*

production: [*underlying form*]$_{UF}$ → [*surface structure*]
comprehension: [*overt form*]$_{OF}$ → { [*surface structure*], [*underlying form*]$_{UF}$ }

We will now explain how Tesar & Smolensky's acquisition model handles these three forms. According to a proposal by Smolensky (1996a), the learner will use for comprehension the same Optimality-Theoretic grammar that she uses in production. Tesar's point is that if we include the overt form in our theory of phonological representations, Smolensky's bidirectionality proposal means that the learner will infer the two hidden structures from the overt form by using her single OT grammar. Thus, if the learner's constraint ranking happens to be as in (1), i.e. trochaic right-aligning, the learner will interpret the overt form [σ σ́ σ σ]$_{OF}$ as the phonological structure [σ (σ́ σ) σ]. The interpretation tableau (25) shows how this works.

(25) *Grammar-guided interpretation by the learner*

[σ σ́ σ σ]$_{OF}$	TROCHAIC	IAMBIC	ALIGNFT-R	ALIGNFT-L
[(σ σ́) σ σ]	*!		**	
☞ [σ (σ́ σ) σ]		*	*	*

Tesar & Smolensky propose that when the listener is confronted with an overt form, she reacts by guessing the adult surface form by taking into account all surface forms that match the given overt form. Thus, when confronted with the overt form [σ σ́ σ σ]_OF, the listener will consider all the surface forms that would give rise to that form, i.e. all possible surface structures that have a single stressed syllable in the middle. Under the same simplifying assumptions as before, there are only two such forms, namely [(σ σ́) σ σ] and [σ (σ́ σ) σ]. The latter will win. Tesar & Smolensky call this process *Robust Interpretive Parsing* (RIP): the learner will assign to the overt form a structure that minimally violates her constraint ranking, even if this structure is ungrammatical in her own production (as it is here, because she would pronounce an underlying [σ σ σ σ]_UF as [σ σ (σ́ σ)]). This makes a large difference for EDCD, which is now given not the correct adult form, but a possibly incorrect form that results from the child's interpretation. Instead of (22), we now have the production tableau (26). Note that a single constraint ranking (i.e. a single grammar) accounts for two 'processes', production and interpretation.

(26) *Learning from the first interpreted input-output pair*

[σ σ σ σ]_UF	TROCHAIC	IAMBIC	ALIGNFT-R	ALIGNFT-L
[(σ́ σ) σ σ]		*	*!*	
[(σ σ́) σ σ]	*!		**	
√ [σ (σ́ σ) σ]		*	*!	*
[σ (σ σ́) σ]	*!		*	*
☞ [σ σ (σ́ σ)]		*		**
[σ σ (σ σ́)]	*!			**

EDCD will now demote ALIGNFT-R below ALIGNFT-L, so that the third candidate becomes better than the fifth. If now, as before, the learner hears the trisyllabic form [σ σ́ σ]_OF, she will interpret this as the non-adultlike form [σ (σ́ σ)], as shown in (27).

(27) *Grammar-guided interpretation by the learner*

[σ σ́ σ]_OF	TROCHAIC	IAMBIC	ALIGNFT-L	ALIGNFT-R
[(σ σ́) σ]	*!			*
☞ [σ (σ́ σ)]		*	*	

The learner can now compute what she herself would have said. This is [(σ́ σ) σ], as shown in tableau (28).

(28) *Learning from the second interpreted input-output pair*

[σ σ σ]$_{UF}$	Trochaic	Iambic	AlignFt-L	AlignFt-R
☞ [(σ́ σ) σ]		*		*
[(σ σ́) σ]	*!			*
√ [σ (σ́ σ)]		*	*!	
[σ (σ σ́)]	*!		*	

By comparing the child's production [(σ́ σ) σ] with her interpretation of the adult form ([σ (σ́ σ)]), EDCD will demote AlignFt-L below AlignFt-R, thereby returning to the situation in (5). This is an example of the general way in which RIP/EDCD can get stuck: because of a non-adultlike interpretation of overt forms, the learner will rerank the wrong constraints and end up visiting an eternal cycle of inadequate grammars, i.e. grammars that produce a non-adultlike overt form for at least one underlying form. In the example at hand, no sequence of adult overt forms [σ σ́ σ]$_{OF}$, [σ σ́ σ σ]$_{OF}$, and [σ σ́ σ σ σ]$_{OF}$ will lead to a correct reranking of the two foot form constraints Trochaic and Iambic; only if the language contains the disyllabic form [σ σ́]$_{OF}$ will the learner be able to rank these constraints correctly and ultimately come up with an adequate grammar.

Neither EDCD nor GLA are guaranteed to lead to successful learning if the learner has to work with three levels of representation. Our example of an iambic left-aligning language turns out to be unlearnable if it has only trisyllabic and quadrisyllabic words. In general, a failure to learn particular hypothetical languages need not be evidence against the learning algorithm; it would even be evidence in *favour* of the learning algorithm if it turned out that the hypothetical language does not actually exist in reality. The current example seems to be such an instance: languages tend to have bisyllabic words as well, and this indeed happens to guarantee learnability in this example, because [σ σ́]$_{OF}$ can only be interpreted as the trochaic form [(σ σ́)]$_{OF}$, which leads to demotion of Trochaic in EDCD. But in a computer simulation of 124 artificial languages (a metrical stress example with 12 constraints), Tesar and Smolensky (2000: chapter 4) found that their combination of Robust Interpretive Parsing with EDCD (RIP/CD) is capable of learning only 60% of these languages, if the constraints start out equally ranked.

As said, the non-optimal performance of RIP/CD is not necessarily evidence against the appropriateness of general constraint-ranking procedures as models of language acquisition. It is not bad in general for a learning algorithm to fail on certain input data. If an OT learning algorithm can predict that certain constraint rankings are unlearnable, and exactly these rankings turn out not to occur in the languages of the world, such an unlearnability result constitutes positive support for that algorithm (cf. Clark & Roberts 1993 for parameter setting). If a failure of RIP/CD predicts a gap in the factorial typology, and if RIP/CD does work for all attested languages, this would actually be evidence in favor of this learning algorithm. Nevertheless, the low success rate of 60 percent has been regarded as problematic, because it is likely that the 40 percent failures contain some languages that actually exist. If the required success rate is 100 percent, RIP/CD must be considered just as bad as general parameter-setting

algorithms. The only working acquisition model for stress systems to date is the *cue-based learning procedure* by Dresher (1999), which regards the acquisition process as a fixed sequence of triggered parameter settings, but since this model is specific to the stress acquisition problem, it requires not only an innate set of parameters, but also an innate learning procedure specific to each parameter. There have been several attempts to improve upon the 60 percent success rate. Tesar & Smolensky themselves try various initial rankings. The success rate rises if in the initial state constraints for iambicity and trochaicity outrank the alignment and other constraints, and it rises again if the weight-to-stress principle is initially ranked even higher, but not to 100 percent. Replacing EDCD with GLA in both the interpretation and production phases raises the success rate as well, but not much (Boersma, to appear).

It is clear that much research is needed to find out whether there is any combination of algorithms, constraint sets, and initial states that accurately predicts the learnability of attested languages and at the same time is capable of showing that many attested gaps in the factorial typology are not accidental but can be explained by the formal unlearnability of such languages. An example of such research outside phonology is Jäger (to appear), who goes even further by predicting not only some learnable and unlearnable languages but also some learnable but diachronically unstable languages.

3.3. Overt and covert 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? There are two kinds of subset problems: overt and covert.

The *overt subset problem* is the problem that a learner may end up in an *overt superset language*, i.e., that she will produce all possible adult forms *and some more*. This problem is related to *optionality* in production. The classical example in morphology is the English-speaking child who maintains two past tenses of the verb *go*, namely the rule-based form *goed* and the adult form *went*. How can the learner decide that she should stop saying *goed*, if explicit negative evidence (an adult telling the child explicitly that *goed* is incorrect) does not work? This problem is solved by the Gradual Learning Algorithm (Boersma 1997, Boersma & Hayes 2001). If the learner hears an adult say *went*, there is a chance that she computes that she herself would have produced *goed*. Noticing the mismatch, the learner will raise the constraint that favours *went*, and lower the constraint that favours *goed* (perhaps a generic *-ed* rule). The reverse case does not occur, since the adult never says *goed*. The learner will thus end up saying *went* all of the time. The learner's own *goed* productions thus constitute the indirect negative evidence necessary for remedying the overt subset problem. 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.

The *covert subset problem* is related to the maxim of *Richness of the Base* (Prince & Smolensky 1993: 191; Smolensky 1996b), according to which 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 exclusively contains forms without codas, but by a grammar that would convert any (perhaps non-existent) lexical form with a coda to a form without, perhaps by means of a high-ranked NOCODA constraint. We call this the *covert* subset problem because languages without codas will typically have lexical forms without codas (as a result of Lexicon Optimization), so that the learner would not generally produce any forms with codas even if NOCODA is ranked low; empirical evidence for a high ranking of coda constraints in languages with restrictions on codas is to be found in processes like loanword adaptation and second-language learning (e.g. English *bed* has been borrowed into Japanese as *beddo*). Generally, grammars will show few covert subset problems if structural (or markedness) 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 desirable property is provided by starting from an initial state in which all structural constraints outrank all faithfulness constraints (M >> F, Smolensky 1996b). But the learning algorithm EDCD by Tesar & Smolensky (1998, 2000) 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.

The Gradual Learning Algorithm again solves EDCD's severe covert subset problems to a large extent. Its reaction to a learning datum that requires a certain structural constraint to 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 could well be comparable to that exhibited by real children. Nevertheless, Hayes (to appear) and Prince & Tesar (1999) propose complicated extensions to EDCD that cause some active enforcement of M>>F rankings throughout the acquisition process. These extensions do not always lead to the most restrictive grammar, and would therefore, like the GLA, predict *some* overly faithful behaviour in loanword adaptation and L2 acquisition.

3.4. Acquiring underlying forms

One of the tasks for the language-learning child is to create abstract underlying forms that can serve as lexical representations and as the inputs to the production grammar. A common assumption in acquisition research is that a young child directly stores perceived forms in her lexicon, without taking into account any morphologically related forms. Thus, the German words that mean 'advice' and 'wheel' will both be stored as [raːt]$_{\text{UF}}$ by a German learning child. At some point, however, this child will

note the relationship of one of them with the word [reːdər]$_{UF}$ 'wheels', and thus posit a new underlying form for 'wheel', namely [raːd]$_{UF}$, which will surface as [raːt] as a result of final devoicing. Tesar & Smolensky (2000: 79) give an Optimality-Theoretic account of this, showing that the [raːd]$_{UF}$ – [reːdər]$_{UF}$ pair is the optimal paradigm. Tableau (29) shows how this works.

(29) *Interpretation of an overt paradigm*

[raːt]$_{OF}$ – [reːdər]$_{OF}$	ONSFAITH	*VOI	FAITH	*V[-VOI]V
☞ raːd$_{\langle voi \rangle}$ – reːdər		*	*	*
raːt – reːt$_{[voi]}$ər	*!	*	*	

The constraint ONSFAITH requires voicing faithfulness for onsets. It is violated in the surface structure [reːt$_{[voi]}$ər], because voicing has been inserted (this is a *containment*-style surface form, as in Prince & Smolensky 1993). The constraint FAITH requires voicing faithfulness in general; it is violated by [raːd$_{\langle voi \rangle}$], which has final devoicing, and of course by [reːt$_{[voi]}$ər]. The constraint *VOI punishes voiced obstruents, as in [reːdər] and [reːt$_{[voi]}$ər]. The constraint *V[-VOI]V punishes intervocalic voiceless consonants. Tableau (29) asserts that it is better to posit an underlying [raːd]$_{UF}$ with final devoicing in the singular than an underlying [raːt]$_{UF}$ with intervocalic voicing in the plural. Of course, the ranking in (29) has to be established separately, perhaps by noting that German routinely allows intervocalic voiceless consonants elsewhere. As in §3.2, therefore, several acquisition tasks have to be performed simultaneously, but Tesar & Smolensky do not yet give a simulation that could show how well their proposed combination of lexical and production learning performs.

An important question is which of the members of a paradigm is taken as the base form for the remaining members. If the phenomenon of *analogical leveling* can be taken as evidence for the spreading of the phonological details from a base form to the other forms of the paradigm, then it appears that the base form for nouns is usually the nominative singular and the base form for verbs is usually a present tense form. However, frequency effects can override this tendency. For instance, the German verb *verliesen* 'lose' turned into *verlieren* by an analogy with the past participle *verloren* 'lost' and in Afrikaans the original past tense *verloor* 'lost' is nowadays the present tense and infinitive of this verb; not accidentally, 'lose' is one the few verbs that occurs much more often in past forms than in present forms. But Albright (2002) proposes that it is often not the most basic or frequent form, but the most informative form that is taken as the base form. One of his examples is the history of the Classical Latin form *honor* 'honor-NOMSG'. The original nominative singular was *honoːs*, and the genitive was *honoːsis*. Rhotacization of intervocalic *s* then led to a genitive form *honoːris* and many other forms with *r* throughout the paradigm. Subsequently, the nominative singular turned into *honoːr* (whence *honor*) by analogy with the other forms, suggesting that one of the oblique forms with intervocalic *r* was used as the base form of the paradigm. Albright proposes a formal learning algorithm, the *minimal generalization learner*, which shows that indeed the oblique forms are more informative in predicting paradigms than the nominative.

3.5. Recognition as a non-trivial mapping from surface to underlying forms

The child not only has to learn abstract underlying forms (§3.4), but also the mapping from overt or surface forms to these. In the example of §3.1 and §3.2, which was about the assignment of metrical structure in languages without lexically determined stressed syllables, both the overt form and the underlying form could be computed easily from the surface structure. For instance, the overt form $[\sigma \; \acute{\sigma} \; \sigma]_{OF}$ is computed from $[(\sigma \; \acute{\sigma}) \; \sigma]$ by removing the hidden structure (the parentheses in this case), and the underlying form $[\sigma \; \sigma \; \sigma]_{UF}$ is computed from $[(\sigma \; \acute{\sigma}) \; \sigma]$ by removing parentheses and stresses. In Tesar & Smolensky's view of what a surface form is, the triviality of the mappings from the surface form to the overt and underlying forms is not a simplifying assumption but a necessity. They call the surface form the *full structural description*, a representation that is understood to contain enough information to derive both the underlying form and the overt form. This notion of *containment*, which was taken from the original formulation of OT by Prince & Smolensky (1993), can be illustrated most clearly by considering faithfulness violations in segmental phonology. Tesar & Smolensky mention the case of the German word *Tag* 'day-NOMSG', whose full structural description could be written $[ta{:}g_{\langle voi \rangle}+\varnothing]$. In this form, the subscript "⟨voi⟩" signals the deletion of the voicing feature (by the German rule of word-final devoicing), and "+∅" stands for the concatenation with a null morph, which is the phonological part of the NOMSG morpheme. The underlying form can be computed trivially from the full structural description by erasing all the addition and deletion signals, i.e. in this case by removing the "⟨voi⟩", which leaves us with $[ta{:}g+\varnothing]_{UF}$. The overt form can be computed trivially from the full structural description by erasing the morphological markers ("+"), erasing the null material ("∅"), and implementing the additions and deletions, in this case only the deletion of "voi", which turns [g] into [k]. This leaves us with the overt form $[ta{:}k]_{OF}$.

But newer versions of Optimality Theory did away with the notion of containment, which Tesar & Smolensky's acquisition model so heavily relied upon. In Correspondence Theory (McCarthy & Prince 1995), surface forms contain no morpheme boundaries, no reference to unpronounced material, and no indication that pronounced material could be absent underlyingly. Thus, the surface form of 'day-NOMSG' in German is simply [ta:k]. Since surface forms no longer contain no information about underlying forms, faithfulness constraints now have to evaluate explicitly the similarity between two separate representations, namely the underlying form and the surface form, as in (30).

(30) *Final devoicing in production*

[ta:g]$_{UF}$ 'day'	*[–son,+voi]/_]$_W$	IDENT(voi)
ta:g	∗!	
☞ ta:k		∗

Faithfulness constraints are not restricted to segmental phonology. In metrical phonology, one often has lexical stress, which comes with constraints like MAX(stress) 'an underlying stress should appear on the surface'. There are also cases

in which structural metrical constraints interact with 'segmental' faithfulness. For example, in (23) we really need high-ranked faithfulness constraints to make sure that the surface form contains three syllables like the underlying form, perhaps a constraint like MAX(σ) 'an underlying syllable should appear on the surface'; otherwise, the candidate [(σ σ́)] would win because it satisfies both foot alignment constraints, as shown in tableau (31).

(31) *Syllable deletion when metrical constraints outrank faithfulness*

[σ σ σ]_UF	ALIGNFT-L	ALIGNFT-R	MAX(σ)	IAMBIC	TROCHAIC
[(σ́ σ) σ]		*!		*	
[(σ σ́) σ]		*!			*
[σ (σ́ σ)]	*!			*	
[σ (σ σ́)]	*!				*
[(σ́ σ)]			*	*!	
☞ [(σ σ́)]			*		*

Exactly this type of optimal faithfulness violations was proposed by Curtin & Zuraw (2002) as an explanation of syllable deletion in child Dutch.

Now why this discussion of faithfulness? The answer is that in all cases where structural constraints outrank faithfulness constraints, the mapping from surface form to underlying form (which can be called *recognition*) becomes non-trivial in a correspondence view of phonological representations. Thus, the surface form [taːk] could map to any of the underlying forms [taːg+∅]_UF, [taːk+∅]_UF, [taːg]_UF, or [taːk]_UF, depending on what is in the lexicon. Tableau (32) shows this.

(32) *Recognition by faithfulness*

[taːk]	IDENT(cont)	IDENT(place)	MAX(ə)	IDENT(voi)	IDENT(a)
[taːt+∅]_UF 'deed'		*!			
☞ [taːg+∅]_UF 'day'				*	
[taːg+ə]_UF 'days'			*!	*	
[park+∅]_UF 'park'		*!			
[zaxə+∅]_UF 'matter'	*!*		*	*	

We see an example here of how GEN restricts the candidate set to the forms that occur in the lexicon. None of the forms satisfies all faithfulness constraints. Note that the candidate [taːg+ə]_UF violates MAX(ə) 'a schwa in the underlying form should occur in the surface form', and the candidate [park+∅]_UF violates IDENT(a) 'preserve the quality of [a]-like vowels'. These constraints are worded as if they refer to production, but they can equally well be used for the reverse mapping, as they are here. The winner is the form [taːg+∅]_UF, which violates only IDENT(voi). This constraint has to be low ranked because this is German, a language in which voiceless

plosives have to be routinely mapped to voiced plosives in the lexicon. Another reason why IDENT(voi) must be low ranked is that it is routinely violated in production, as in (30). Of course, the two reasons are strongly related since all faithfulness violations that the speaker produces will have to be undone by the listener. As a further illustration, we also gave a low ranking to IDENT(a), to emphasize the fact that /ar/ and /aː/ are almost homophonous, i.e., the surface forms [taːk] and [park] are pronounced approximately as [tʰaːk]_OF and [pʰɑːk]_OF.

The idea that faithfulness can handle the mapping from surface to underlying form was defended by Smolensky (1996a), but criticized by Hale & Reiss (1998), who argued that it would mean that the German surface form [ʀaːt] would always be recognized as [ʀaːt+∅]_UF 'advice', never as [ʀaːd+∅]_UF 'wheel', as shown in tableau (33).

(33) *Failure to recognize 'wheel' if there are only faithfulness constraints*

[ʀaːt]	IDENT(cont)	IDENT(place)	MAX(ə)	IDENT(voi)	IDENT(a)
☞ [ʀaːt+∅]_UF 'advice'					
[ʀaːd+∅]_UF 'wheel'				*!	
[taːt+∅]_UF 'deed'	*!	*		*	

The winning candidate simply violates no faithfulness constraints at all, but it is clear that in a semantic context that involves bicycles the candidate [ʀaːd+∅]_UF 'wheel' should sometimes win. To tackle this situation, we need constraints that evaluate the likelihood of any lexical item in the given semantic context (Boersma 2001). Since these anti-lexical-access constraints evaluate underlying forms only, they can be called by the name *LEX. The following example is from Boersma (2001). In the case of final devoicing in Dutch, the words [ʀɑd]_UF 'wheel' and [ʀɑt]_UF 'rat' are both pronounced [ʀɑt]_OF. For simplicity, we assume that the surface form is [ʀɑt] in both cases, i.e. that devoicing has to be accounted for by the phonology. In the semantic context 'turn', the listener has to map [ʀɑt] to [ʀɑd]_UF 'wheel'. This mapping involves a replacement of a [–voi] feature value by [+voi], and it seems to be the case that we need to model this with an OT grammar with low-ranked faithfulness constraints again to ensure that the listener will allow this mapping. This grammar also needs high-ranked faithfulness constraints in order to make sure that the listener does not map [ʀɑt] to [ʋil]_UF 'wheel'! The full *recognition* procedure is shown in tableau (34).

(34) *Recognizing 'wheel'*

[ʀɑt] context = 'turn'	*LEX([ʀɑt]UF 'rat' /'turn')	IDENT (height)	*LEX ([ʀɑd]UF 'wheel' /'turn')	*LEX ([ʋil]UF 'wheel' /'turn')	IDENT (voi)
[ʀɑt]UF 'rat'	*!				
☞ [ʀɑd]UF 'wheel'			*		*
[ʋil]UF 'wheel'		*!		*	*

Given the perceived surface form [ʀɑt] and the semantic context 'turn', a high-ranked *LEX constraint will militate against recognizing the surface form as the word meaning 'rat'. As far as the semantics is concerned, tableau (34) shows that the lexicon favours recognition of the word [ʋil]UF 'wheel', since that word is appropriate in the context and it is much more common in Dutch than its near-synonym [ʀɑd]UF 'wheel'. But a high-ranked faithfulness constraint against replacing /ɑ/ with /i/ will prevent recognition of [ʋil]UF 'wheel'. The remaining candidate is [ʀɑd]UF 'wheel', which is neither semantically nor phonologically far off from the input.

Boersma (2001) shows that the GLA can account for the acquisition of some attested phenomena in recognition, namely the dependence on frequency (*LEX comes to be lower ranked for [ʋil]UF than for the less common [ʀɑd]UF), and the dependence of the ranking of *LEX on the sematic context itself. Boersma's account of recognition is probably overly simplistic, because many more things than semantic contexts have to be taken into account, let's say the whole of syntax and pragmatics. For this good reason, modelling recognition and its acquisition is not the favourite subject of most phonologists. This is where neat modular views of phonology start to leak.

3.6. Perception as a non-trivial mapping from surface to underlying forms

In Tesar & Smolensky's metrical example, the mapping from overt to surface form (which we can call *perception*) used the same constraints as production. But in general this mapping may have its own specific constraints. One has to realize that the overt form of *Tag* includes language-specific phonetic detail, for instance aspiration, as in [tʰaːk]OF. The question, now, is whether the surface form has this phonetic detail as well, i.e. whether the surface form is [tʰaːk] or [taːk]. But for faithfulness constraints to be able to evaluate the similarity between surface form and underlying form, the surface form has to be *commensurable* with the underlying form, which means that if the underlying form consists of arbitrary symbols taken from a small set (for reasons of lexical economy), the surface form will have to consist of the same type of arbitrary symbols and there is no room in the surface form for phonetic detail. Thus, the surface form may have to be [taːk]. In interpretation, the mapping from the aspirated phone [tʰ]OF to the abstract segment /t/ (this mapping can be called *perception*) is specific to German, and may therefore have to be handled by a perception grammar that contains constraints that relate phonetic detail (auditory

cues) to abstract phonological representations. The learnability of such mappings is discussed by Escudero & Boersma (2003).

3.7. Emergence of constraints

As alluded to in §2.3, not all constraints are universal, and many cannot be innate, because they refer to material that is specific to the language at hand, especially to specific morphemes. Albright & Hayes (2003) discuss the emergence of morphological constraints from alternations that the child hears in her environment. But even purely phonological constraints could be selected from a larger pool: Hayes (1999) proposes a process called *inductive grounding* to create an initial state of the grammar that consists only of *phonetically grounded*, but formally simple, constraints. For instance, Hayes proposes that the constraint *[Lab,–voi], which militates against voiceless labials (like the /p/ that is lacking in Arabic), enters the child's initial grammar because the child can calculate that this constraint is a better predictor of articulatory effort than its formally equally simple or simpler neighbours *[Cor,–voi], *[Dors,–voi], *[Lab,+voi], *[Lab], and *[–voi]. It is even possible to think that *all* phonological constraints emerge during acquisition (Boersma 1998): faithfulness constraints for a phonological category are created as soon as the infant has created that category, and articulatory constraints against an articulatory gesture are created as soon as the child has learned how to use that gesture to implement a phonological category; such a view of the grammar lacks markedness constraints, because markedness is considered an epiphenomenon of the rankings of both faithfulness and articulatory constraints. The first step in this sequence, namely the creation of phonological categories, can be based on non-lexically-driven distributional learning, which can be modeled within OT, as has been shown by Boersma, Escudero & Hayes (2003).

4. How should the field proceed?

We have shown how the basic tenets of OT can be applied to empirical data and how OT fares as a learnable theory. The picture we presented of the accomplishments is, on the whole, positive and optimistic – even though we are aware of less felicitous applications. No previous phonological theory had the combined merits for acquisition research of (1) a formal structure, (2) markedness being embedded in this formal structure, (3) the possibility to view the adult grammar as a continuation of the initial grammar, i.e. the continuity hypothesis, (4) a single grammar controlling segmental and prosodic structure, as well as their interactions, (5) the possibility to account for variation and variable forms in a non-stipulative, formal way, and (6) having the properties of a learnable theory. This sparks off our positive attitude at this stage.

We will now turn to the future, and discuss how the field could or should proceed. The respondents to our survey show a remarkable agreement on the topics that should be addressed in the future, namely (1) the relation between computational models of learnability and empirical studies of acquisition, and (2) the relation

between perception and production. Initial steps have been taken in these directions, and we will discuss some of this work, just to leave the reader with a flavor of the great opportunities that lie ahead.

4.1. Learning algorithms and empirical data

In §3.1 we referred to the computer simulation of syllable type acquisition carried out by Boersma & Levelt (2000). The Gradual Learning Algorithm (GLA) had to acquire the ranking for Dutch of the syllabic markedness constraints discussed in §2 vis-à-vis a faithfulness constraint FAITH. In the GLA constraints have ranking values, and the initial M >> F state of the grammar was modelled by arbitrarily setting the ranking value of the markedness constraints to 100 and that of FAITH to 50. The input to the learning algorithm consisted of thousands of syllables, in the distribution found for Dutch. The result of the simulation is in Figure 1.
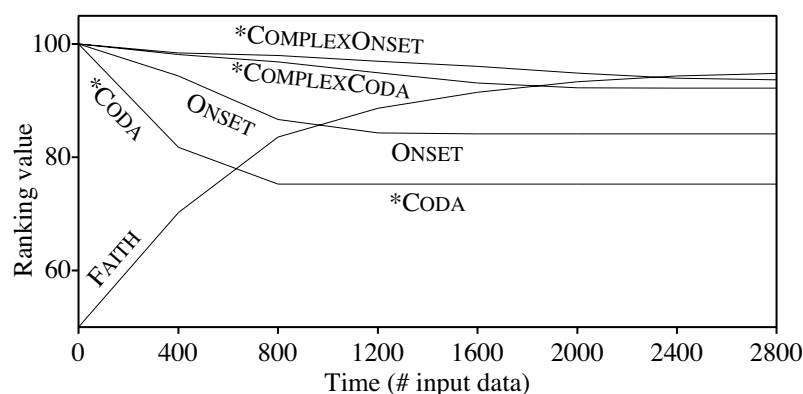

Fig. 1. Developing constraint rankings (from Boersma & Levelt 2000).

As can be seen in the figure, FAITH gradually rises in the ranking, while the markedness constraints gradually descend. In this process, FAITH first overtakes *CODA (NoCODA), then ONSET, and finally *COMPLEX-O and *COMPLEX-C. This exactly replicates the attested order of acquisition. Less simplistic GLA learning curves can be found in Curtin & Zuraw (2001) and Jäger (to appear).

The GLA is also capable of modeling variability in production. In empirical data variable productions occur often, reflecting a transition between two grammatical stages (Adam 2003). In a single recording session a child may produce a word like *cat* both as [kæ], without a coda, and as [kæt], with a coda. In GLA this occurs when NoCODA and FAITH have come to lie close to each other at some point during the reranking process. Because of *noisy evaluation* it will be impossible to come up with a categorical ranking of constraints that lie close together. The grammar will therefore be unstable and alternate between a NoCODA >> FAITH ranking and a FAITH >> NoCODA ranking at evaluation time, resulting in variable productions.

Combining computational models with empirical data opens up a whole new area of research, in which all kinds of hypotheses can be tested in a controlled way. For one thing, it provides a great opportunity to test hypotheses about the role that the distribution of forms in the surrounding language plays in development. The potential effect of frequency on phonological development has been indicated in several

empirical studies (Roark & Demuth 2000; Maye & Gerken 2002; Fikkert & Levelt 2002; Fikkert, Levelt & Van de Weijer 2003; Levelt & Van de Vijver, to appear) but needs more definitive confirmation. In addition, combining computational models and empirical studies can help to get more insight in the nature of variable productions: when do they occur, and in what ways can productions vary at a given moment? This would be a welcome supplement to studies on variation in child language (Gierut, Morrisette & Champion 1999; Dinnsen & McGarrity 1999). A computational model can make predictions that can be tested in empirical data. Vice versa, computational models can be improved by taking into account empirical data.

4.2. Perception and production

Interdisciplinary research on comprehension and production has various dimensions. The relation is obvious: the perceived input needs to be mapped onto an underlying representation, which in turn forms the input for output candidates that will be produced. We saw that Tesar & Smolensky distinguish three representations (overt, surface, and underlying) that are related in such a way that the same grammar that evaluates output candidates in production also co-determines the mappings from overt to surface form (perception) and from surface to underlying form (recognition). However, their metrical examples that we discussed in §3 simplified the two comprehensive mappings in such a way that faithfulness constraints were not needed (there was always the same number of syllables in all three forms), whereas more comprehensive grammar models that include segmental phonology probably require faithfulness constraints that are specific to one or both of the comprehension mappings. We will discuss three approaches to this issue.

4.2.1. Separate grammars for perception and production

The mapping from the raw overt form to an abstract phonological surface structure involves the mapping of phonetic details to phonological categories. Boersma (1998) models this process with constraints that are specific to categorization. Since the result of this mapping is a phonological structure, many of the same structural constraints that are often considered part of the production grammar play a role in the perception grammar as well. Boersma takes the strong point that if these structural constraints reside in the perception grammar, their role in the production grammar is greatly reduced; Boersma proposes that faithfulness constraints in production evaluate the extent to which the speaker thinks the listener will be able to reconstruct the phonological structure without lexical access. In other words, faithfulness equals recoverability. The model is summarized in (35).

(35) *Boersma's model of phonology*

perception: [*overt form*]$_{OF}$ → [*surface structure*]
recognition: [*surface structure*] → [*underlying form*]$_{UF}$
production: [*underlying form*]$_{UF}$ → { [*overt form*]$_{OF}$ → [*surface structure*] }

This model allows the inclusion of phonetic detail in the grammar, and therefore the direct evaluation of functional principles like minimization of perceptual confusion

and articulatory effort. Like Tesar & Smolensky's model discussed in §3.5, the acquisition of Boersma's model involves the concurrent acquisition of three grammars: perception (Boersma 1997), production (Boersma & Hayes 2001), and recognition (Boersma 2001, discussed above in §3.5). No simulations of this triple acquisition have been provided yet.

4.2.2. Perception and production in a single grammar

In infant language development, receptive competence seems to precede productive competence. There is, therefore, a developmental gap between comprehension and production. Smolensky (1996a) showed that a single grammar can capture both perfect comprehension – a perceived adult output is paired with a target appropriate input – and immature production. However, research by Stager and Werker (1997) indicated that in word-learning tasks, 14-months old children heard less phonetic detail than in a discrimination task. Acoustic information can thus be encoded at the phonetic level of representation, while encoding at the lexical level of representation fails. Pater (to appear) therefore concludes that there are actually two gaps: one between the child's phonetic representation and the child's lexical representation, and one between the child's lexical representation and the child's output. Pater further observes that the lexical representation is, with a few exceptions, target-appropriate by the time production starts. The developmental gap between the phonetic and lexical representations thus precedes the gap between comprehension and production. The interesting question is whether the two pairs of representations, phonetic/lexical and lexical/output, are moderated by the same set of constraints, at different times. Early productions are heavily affected by markedness constraints: are early lexical representations affected by these same markedness constraints? There are some experimental results that point in this direction. For example, infants were shown to listen longer to a text containing familiarized words consisting of a single trochaic foot, than to a text containing familiarized words with an iambic foot shape (Jusczyk, Houston & Newsome 1999). This could mean that the same constraints that are operative at the stage where longer adult targets are truncated, discussed in §2.1, are operative somewhere in the mapping from a phonetic to a lexical representation.

Pater (to appear) considers several grammar models. In the most elaborate of these, he entertains the same three representations as Tesar and Boersma (overt, surface, lexical) and the same faithfulness constraints as Boersma (overt-to-surface, surface-to-lexical, and lexical-to-surface), but hypothesizes (with Tesar) that all constraints are in a single grammar so that the same structural constraints can evaluate the surface form in perception as well as in production; moreover, these structural constraints are capable of evaluating the lexical form in recognition. Four possible rankings then correspond to the four different developmental stages in (36).

(36) *Developmental grammars for truncation*

Stage 1: **STRUCT** >> { FAITH(o-s), FAITH(s-l), FAITH(l-s) }
⇒ truncation in perception (hence in the lexicon) as well as in production.
Stage 2: FAITH(o-s) >> **STRUCT** >> { FAITH(s-l), FAITH(l-s) }
⇒ truncation in production and in the lexicon but not in perception.

Stage 3: { FAITH(o-s), FAITH(s-l) } >> **STRUCT** >> FAITH(l-s)
    ⇒ truncation in production but not in perception or in the lexicon.
Stage 4: { FAITH(o-s), FAITH(s-l), FAITH(l-s) } >> **STRUCT**
    ⇒ no truncation.

We thus see some theoretical convergence in the sense that authors agree that theoretical acquisition models should entertain at least three representations, and that the mappings between these should be handled by Optimality Theory. There is no general agreement yet on the number of faithfulness families (are FAITH(s-l) and FAITH(l-s) the same?), on whether the same structural constraints evaluate the outputs of perception and production, on whether structural constraints can evaluate lexical representations, and of course on what the precise formulation of the constraints is and where they come from.

4.2.3. Experimental approaches

Experimental approaches are of course related to the discussion in §4.2.2. The idea here is to research the effect of well-established markedness constraints on the linguistic processing of infants. An experimental research program has been proposed by Smolensky, Davidson & Jusczyk (to appear), in which OT principles concerning the competence grammar are linked to the performance of infants by a set of Linking Hypotheses. The basic hypothesis is that infants will attend longer to stimuli that conform better to their current grammar, all else being equal. Initial results are promising for this hypothesis. It turned out, for example, that young infants (under 6 months), which are hypothesized to have a M >> F ranking of constraints, had longer listening times for CV syllables – which conform to this state of the grammar – than for VC syllables – which violate the high-ranked markedness constraints.

Others have set out to establish experimentally the target-appropriateness of the underlying representation (Nijmegen/Utrecht research group on the feature [voice] of Fikkert and Kager). The results of the Stager and Werker study indicate that lexical representations, i.e. underlying representations, are unspecified in certain respects. It is claimed that the underlying representation is similar to the adult surface representation by the time children start to produce meaningful speech (see the discussion of Pater's work in §4.2.2) but this has not been verified. Certain child language data, like the famous *puzzle-puddle-puggle* case of Smith (1973), discussed in Macken (1980) and in the OT accounts of Boersma (1998) and Dinnsen, O'Connor & Gierut (2001), or the holistic word-forms discussed in Waterson (1971), Levelt (1994) and Fikkert & Levelt (2002) appear to result from the combination of an unspecified representation and an immature grammar, rather than from the grammar alone. The real art is to obtain both production data and perception data of children in a study, and to combine these results in such a way that the nature of lexical representations at a particular stage of production can be established.

Many more directions in which the field could proceed can be imagined. The interaction of morphology and phonology in acquisition will definitely be tackled, just like interactions between prosodic and segmental structure. Finally, while there is a

general idea about an initial grammar where M >> F, initial or early rankings within the two sets of constraints forms another area to be explored.

5. Conclusions

The introduction of Optimality Theory has stirred up the field of phonological acquisition. We hope that our review, albeit selective, has nevertheless illuminated the ways in which Optimality Theory can be, and has been, applied to acquisition research.

What is still missing is generally accessible empirical data. Ideally, scientific research is replicable; people should be able to check proposed theories and explanations against the data that these proposals claim to account for. Yvan Rose (Rose 2003) has designed an easy-to-handle database, *ChildPhon*, available through CHILDES (MacWhinney & Snow 1985), specifically designed for phonetically transcribed data, which can be linked to audio files. ChildPhon is meant to form the phonological complement of the CHAT files in the CHILDES database. We hope that many researchers will seize this opportunity to make their data available.

Acknowledgements

References

(ROA = Rutgers Optimality Archive, http://roa.rutgers.edu)

Adam, G. (2003). *From variable to optimal grammar: Evidence from language acquisition and language change*. PhD dissertation. Tel Aviv: Tel-Aviv University.
Albright, A. (2002). *The identification of bases in morphological paradigms*. PhD dissertation, UCLA.
Albright, A. & Hayes, B. (2003). Learning nonlocal environments. Handout LSA meeting, Atlanta.
Barlow, J. (1997). *A constraint-based account of syllable onsets: Evidence from developing systems*. PhD dissertation. Bloomington: Indiana University.
Barlow, J. & Pater, J. (to appear). Constraint conflict in cluster reduction. *Journal of Child Language*.
Bernhardt, B. & Stemberger, J. (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego, CA: Academic Press.
Blevins, J. (1995). The syllable in phonological theory. In J.A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Cambridge (MA) and Oxford: Blackwell.
Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences*, 21, 43–58. University of Amsterdam.
Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam. The Hague: Holland Academic Graphics. [http://www.fon.hum.uva.nl/paul/]
Boersma, P. (2001). Phonology-semantics interaction in OT, and its acquisition. In R. Kirchner, W. Wikeley & J. Pater (Eds.), *Papers in Experimental and Theoretical Linguistics*, Vol. 6 (pp. 24–35). Edmonton: University of Alberta. [ROA 369, 1999]

Boersma, P. (to appear). Review of Tesar & Smolensky (2000). *Phonology*.

Boersma, P., Escudero, P. & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1013–1016). [ROA 585]

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

Boersma, P. & Levelt, C. (2000). Gradual constraint-ranking learning algorithm predicts acquisition order. In E.V. Clark (Ed.), *Proceedings of the 30th Child Language Research Forum* (pp. 229–237). Stanford, CA: CSLI. [ROA 361, 1999]

Clark, R. & Roberts, I. (1993). A computational model of language learnability and language change. *Linguistic Inquiry*, 24, 299–345.

Curtin, S. & Zuraw, K. (2002). Explaining constraint demotion in a developing system. In B. Skarabela, S. Fish, and A.H.-J. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development* (pp. 118–129). Somerville, MA: Cascadilla Press.

Demuth, K. (1995a). Markedness and the development of prosodic structure. In J. Beckman (Ed.), *Proceedings of the North East Linguistic Society 25* (pp. 13–25). Amherst, MA: GSLA.

Demuth, K. (1995b). Stages in the acquisition of prosodic structure. In E. Clark (Ed.), *Proceedings of the 27th Child Language Research Forum* (pp. 39–48). Stanford, CA: CSLI.

Demuth, K. (1996). Alignment, stress and parsing in early phonological words. In B. Bernhardt, J. Gilbert & D. Ingram (Eds.), *Proceedings of the UBC International Conference on Phonological Acquisition* (pp. 113–124). Somerville, MA: Cascadilla Press.

Dinnsen, D.A., & McGarrity, L. (1999). Variation in emerging faithfulness in phonological acquisition. In A. Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of the 23rd annual Boston University Conference on Language Development* (pp. 172–183). Somerville, MA: Cascadilla Press.

Dinnsen, D. & O'Connor, K. (2001). Typological predictions in developmental phonology. *Journal of Child Language*, 28, 597–628.

Dinnsen, D., O'Connor, K. & Gierut, J. (2001). The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition, *Journal of Linguistics*, 37, 503–525.

Dresher, B.E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic Inquiry*, 30, 27–67.

Dresher, B.E. & Kaye, J. (1990). A computational learning model for metrical theory. *Cognition*, 34, 134–195.

Eimas, P., Siqueland, E., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303–306.

Escudero, P. & Boersma, P. (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In S. Arunachalam, E. Kaiser & A. Williams (Eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8.1 (pp. 71–85). [ROA 439, 2001]

Fikkert, P. (1994). *On the acquisition of prosodic structure*. HIL Dissertations in Linguistics 6. The Hague: Holland Academic Graphics.

Fikkert, P. & Levelt, C. (2002). Putting Place into place. Paper presented at the GLOW workshop on the acquisition of phonology, April 13, 2002, Utrecht.

Fikkert, P., Levelt, C. & Van de Weijer, Joost (2003). The infuence of Child Directed Speech (CDS) on the nature of Consonant Harmony. Paper presented at GALA 5, Utrecht.

Gibson, E. & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, 25, 407–454.

Gierut, J., Morrisette, M. & Champion, A. (1999). Lexical constraints in phonological acquisition. *Journal of Child Language, 26,* 261–294.

Gilbers, D. (2001). Conflicting phonologically-based and phonetically-based constraints in the analysis of liquid-nasal substitutions. *Clinical Linguistics and Phonetics*, 15.1/2, 23–28.

Gilbers, D. & Van der Linde, K. (1999). On the acquisition of segments in Optimality Theory: the acquisition of segments as a conflict between correspondence and markedness constraints. In M. Beers & S. Peters (Eds.), *Phonological development: Different perspectives. Antwerp Papers in Linguistics* 96 (pp. 33–68).

Gnanadesikan, A. (1995/to appear). Markedness and faithfulness constraints in child phonology. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Fixing priorities: Constraints in phonological acquisition*. Cambridge: Cambridge University Press. [ROA 67, 1995]

Goad, H. (1997). Consonant harmony in child language: An optimality theoretic account. In S.J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 113–142). Amsterdam: John Benjamins.

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

Hallé, P. & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infant's early receptive lexicon. *Infant Behavior and Development*, 19, 463–481.

Hansson, G. (2000). *Theoretical and typological issues in consonant harmony*. PhD dissertation. Berkeley: University of California.

Hayes, B. (1999). Phonetically-driven phonology: The role of Optimality Theory and Inductive Grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer & K. Wheatley (Eds.), *Functionalism and Formalism in Linguistics*, Vol. I: *General Papers* (pp. 243–285). Amsterdam: John Benjamins. [ROA 158, 1996]

Hayes, B. (to appear). Phonological acquisition in OT: The early stages. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Fixing priorities: Constraints in phonological acquisition.* Cambridge: Cambridge University Press. [ROA 327, 1999]

Inkelas, S. & Rose, Y. (2003). Velar fronting revisited. In B. Beachley, A. Brown & F. Conlin (Eds.), *Proceedings of the 27th annual Boston University Conference on Language Development* (pp. 334–345). Somerville, MA: Cascadilla Press.

Jäger, G. (to appear). Learning constraint sub-hierarchies: The Bidirectional Gradual Learning Algorithm. In H. Zeevat & R. Blutner (Eds.), *Optimality Theory and pragmatics*. Palgrave Macmillan. [ROA 544, 2002]

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala. Translated into English (1968) as *Child language, aphasia and phonological universals*. The Hague: Mouton.

Jongstra, W. (2003). *Variation in reduction strategies of Dutch word-initial consonant clusters*. PhD dissertation. Toronto: Toronto University Working Papers.

Joppen, S. & Grijzenhout, J. (1999). First steps in the acquisition of German phonology: A case study. [ROA 304].

Joppen, S. & Grijzenhout, J. (2000). First steps in the acquisition of German consonants: Minimal constraint demotion. [ROA 400].

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P., Houston, D. & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.

Kiparsky P. & Menn, L. (1977). On the acquisition of phonology. In J. Macnamara (Ed.), *Language learning and thought* (pp. 47–78). New York: Academic Press.

Kirk, C. & Demuth, K. (2003). Onset/coda asymmetries in the acquisition of clusters. In B. Beachley, A. Brown & F. Conlin (Eds.), *Proceedings of the 27th annual Boston University Conference on Language Development* (pp. 437–448). Somerville, MA: Cascadilla Press.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93–107.

Levelt, C. (1994). *On the acquisition of Place*. HIL Dissertations in Linguistics 8. The Hague: Holland Academic Graphics.

Levelt, C. (1995). Unfaithful kids: Place of Articulation patterns in early child language. Paper presented at the University of Pennsylvania, October 13, 1995.

Levelt, C., Schiller, N. & Levelt, W. (2000). The acquisition of syllable types. *Language Acquisition,* 8.3, 237–264.

Levelt, C. & Van de Vijver, R. (to appear). Syllable types in cross-linguistic and developmental grammars. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Fixing priorities: Constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Lléo, C. (2001). The interface of phonology and morphology: The emergence of the article in early acquisition of Spanish and German. In J. Weissenborn & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, syntactic and neurophysiological aspects of early language acquisition*. Amsterdam: John Benjamins.

Lleó, C. & Demuth, K. (1999). Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages. In A. Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of the 23rd annual Boston University Conference on Language Development* (pp. 407-418). Somerville, MA: Cascadilla Press.

Macken, M. (1980). The child's lexical representation: the 'puzzle–puddle–pickle' evidence. *Journal of Linguistics*, 16, 1–17.

MacWinney, B. & Snow, C. (1985). The Child Language Exchange System. *Journal of Child Language*, 12, 271–296.

Maye, J. & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. Fish & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual Boston University Conference on Language Development* (pp. 522–533). Somerville, MA: Cascadilla Press.

McCarthy, J. & Prince, A. (1995). Faithfulness and reduplicative identity. In J. Beckman, L. Walsh Dickey & S. Urbanczyk (Eds.), *Papers in Optimality Theory*. University of Massachusetts Occasional Papers 18 (pp. 249–384). Amherst, MA: Graduate Linguistic Student Association. [ROA 60]

Ota, M. (1999). *Phonological theory and the acquisition of prosodic structure: Evidence from child Japanese*. PhD Dissertation. Washington DC: Georgetown University.

Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition*, 6, 201–253.

Pater, J. (to appear). Bridging the gap between perception and production with minimally violable constraints. In R. Kager, J. Pater & W. Zonneveld (Eds.*)*, *Fixing priorities: Constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Pater, J. & Werle, A. (2001), Typology and variation in child consonant harmony. In C. Féry, A. Dubach Green & R. Van de Vijver (Eds.), *Proceedings of the 5th HIL Phonology Conference* (pp. 119–139). Potsdam: University of Potsdam.

Prince, A. & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report TR-2, Rutgers University Center for Cognitive Science.

Prince, A. & Tesar, B. (1999). *Learning phonotactic distributions*. Technical Report TR-54, Rutgers University Center for Cognitive Science. [ROA 353]

Roark, B. & Demuth, K. (2000). Prosodic constraints and the learner's environment: A corpus study. In S. C. Howell, S. Fish & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual Boston University Conference on Language Development* (pp. 597–608). Somerville, MA: Cascadilla Press.

Rose, Y. (2000). *Headedness and prosodic licensing in the L1 acquisition of phonology.* PhD dissertation. Montreal: McGill University.

Rose, Y. (2001). Licensing and feature interaction processes in child language. In K. Megerdoomian & L. Bar-el (Eds.), *Proceedings of the 20th West Coast Conference on Formal Linguistics* (pp. 484-497). Somerville, MA: Cascadilla Press.

Rose, Y. (2003). ChildPhon: A database solution for the study of child phonology. In B. Beachley, A. Brown & F. Conlin (Eds.), *Proceedings of the 27th annual Boston University Conference on Language Development* (pp. 674–685). Somerville, MA: Cascadilla Press.

Smith, N.V. (1973). *The acquisition of phonology*. Cambridge: Cambridge University Press.

Smolensky, P. (1993). Harmony, markedness, and phonological activity. Handout Rutgers Optimality Workskop 1. [ROA 87]

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

Smolensky, P. (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]

Smolensky, P., Davidson, L. & Jusczyk, P. (to appear). The initial and final states: Theoretical implications and experimental explorations. In R. Kager, J. Pater & W. Zonneveld (Eds.)*, Fixing priorities: Constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Stager, C. & Werker, J. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381-382.

Stampe, D. (1969). The acquisition of phonemic representation. In R. Binnick, A. Davidson, G. Green & J. Morgan (Eds.), *Papers from CLS 5* (pp. 433–444). Chicago: Chicago Linguistic Society.

Tesar, B. (1995). *Computational Optimality Theory*. PhD dissertation, University of Colorado. [ROA 90]

Tesar, B. (1997). An iterative strategy for learning metrical stress in Optimality Theory. In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the 21st annual Boston University Conference on Language Development* (pp. 615–626). Somerville, MA: Cascadilla Press. [ROA 177]

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

Tesar, B. (1999). Robust interpretive parsing in metrical stress theory. In K. Shahin, S. Blake & E.-S. Kim (Eds.), *Proceedings of the 17th West Coast Conference on Formal Linguistics* (pp. 625–639). Stanford, CA: CSLI. [ROA 262]

Tesar, B. (2000). On the roles of optimality and strict domination in language learning. In J. Dekkers, F. van der Leeuw & Jeroen van de Weijer (Eds.), *Optimality Theory: Phonology, syntax, and acquisition* (pp. 592–620). New York: Oxford University Press.

Tesar, B. & Smolensky, P. (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, B. & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, 29, 229–268.

Tesar, B. & Smolensky, P. (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Waterson, N. (1971). Child phonology, a prosodic view. *Journal of Linguistics*, 7, 179–211.

Werker, J. & Stager, C. (2000). Developmental changes in infant speech perception and early word learning: Is there a link? In J. Pierrehumbert & M. Broe (Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon* (pp. 181–193). Cambridge: Cambridge University Press.