Language acquisition and language change in bidirectional Optimality Theory[1]

Petra Hendriks & Jacolien van Rij
University of Groningen

Abstract

Many theories of language assume that the change of a language over time and the acquisition of a language by an individual are interrelated. In this paper we compare and try to relate Mattausch's (2004) diachronic bidirectional Optimality Theory (bi-OT) account of the development of pronominal binding in English to Hendriks and Spenader's (2004; 2005/6) synchronic bi-OT account of English-speaking children's acquisition of pronominal binding. Our examination shows that Mattausch's frequency-based approach does not yield an adequate explanation for children's acquisition and adults' processing of pronominal binding in Modern English. On the other hand, the grammaticalization of Principle B in English is not readily explainable from Hendriks and Spenader's non-frequentist approach to language acquisition. This suggests that a linguistic theory aiming to explain language change as well as language acquisition should take into account both statistical patterns in the language input and linguistic and cognitive factors involved in language use.

1. Introduction

A language is not a fixed system of forms and meanings. Rather, all living languages are constantly changing. This is obvious when we compare the contemporary variant of a language to older variants of the same language. We can also observe language

[1] This investigation was supported in part by a grant from the Netherlands Organisation for Scientific Research, NWO (grant no. 277-70-005 for Petra Hendriks). We thank the three anonymous reviewers for this volume for their useful comments and suggestions.

change when we look at a single individual trying to learn a language. At first, the forms the language learner produces and the meanings the language user understands will be quite different from the adult system of forms and meanings. But with experience, the language produced by the language learner may approach the adult system of forms and meanings closer and closer until mature linguistic competence is achieved. The relation between language acquisition and language change has been made explicit by those assuming that certain types of language change result from a misalignment of the child's analysis and an adult's analysis of the same data (e.g., Lightfoot, 1999). Also, most computational models of language change incorporate a model of language acquisition into them (Pearl and Weinberg, 2007), although sometimes implicitly.

If language change is indeed driven by considerations arising from language acquisition, it is expected that an adequate theory of language change not only explains the process of language change but also yields correct predictions with respect to the acquisition of the language. Conversely, a theory of language acquisition would be strengthened if it could be shown to also account for the way the adult language came to be the way it is. In this paper, we compare Mattausch's (2004) diachronic account of the development of pronominal binding in English with Hendriks and Spenader's (2004; 2005/6) synchronic account of English-speaking children's development of pronominal binding. The two accounts appear to be highly similar in several respects, as they are both couched within the framework of bidirectional Optimality Theory (OT) and they both present an explanation of the adult pattern of pronominal forms and meanings in modern English. Furthermore, Mattausch's model of diachronic change includes a learning component. It is therefore expected that this model is also able to account for children's acquisition of pronominal binding. If not, the fact that the two models are based on similar assumptions suggests that it should at least be possible to combine them into a single model of grammar that is able to account for the entire set of data.

The aim of this paper is to investigate whether it is possible to arrive at a bidirectional OT model that is able to account for the diachronic as well as synchronic development of pronominal binding by combining aspects of the two models. In section 2, we briefly discuss the basic properties of the linguistic phenomenon under discussion, pronominal binding. Section 3 presents the bidirectional OT model of children's development of pronominal binding proposed by Hendriks and Spenader

(2004; 2005/6). Section 4 presents Mattausch's (2004) bidirectional OT model of the evolution of pronominal binding in English. In section 5, the two models are compared and their compatibility is investigated. Section 6 presents our conclusions.

2. Pronouns and reflexives

Modern Standard English distinguishes between pronouns and reflexives, which are more or less in complementary distribution (except in syntactic environments such as locative PPs). Reflexives such as *himself* must be locally bound, whereas pronouns such as *him* cannot be locally bound. This is generally formulated in terms of the two complementary principles A and B of Binding Theory (cf. Chomsky, 1981):

(1) Principle A: Reflexives must be bound locally.

(2) Principle B: Pronouns must be free locally.

An element is locally bound if it is coreferential with a c-commanding potential antecedent within the same local domain. Principle A explains why *himself* in (3) must be coreferential with the local subject *Bert* and cannot refer to someone else.

(3) Bert washed himself.

(4) Bert washed him.

In (4), in contrast, *him* cannot be coreferential with the local subject *Bert* and must refer to someone else. This behavior of the pronoun *him* is explained by Principle B. Although Principle A and Principle B appear to be two highly similar principles, in the next section we will see that children treat these two principles differently.

3. A bi-OT model of language acquisition

In this section, we discuss Hendriks and Spenader's model of language acquisition, which accounts for the well-known asymmetry between children's acquisition of Principle A and their acquisition of Principle B.

3.1 The Delay of Principle B Effect in language acquisition

A well-known observation with respect to the acquisition of pronominal binding is the observation that children acquire the correct meaning of reflexives much earlier than they acquire the correct meaning of pronouns (e.g., Chien and Wexler, 1990; Grimshaw and Rosen, 1990). Presented in a context with two male referents, say Bert and Ernie, sentences like (3) are correctly understood from a young age on. However, children frequently misinterpret *him* in (4) as coreferring with the subject until roughly the age of 6. According to these children, sentence (4) can also mean that Bert washed himself. This delay in the correct interpretation of pronouns is often referred to as the "Delay of Principle B Effect". Interestingly, this delay is only observed in interpretation. The same children's production of pronouns, as well as their production of reflexives, is adult-like from at least the age of 4 on (De Villiers, Cahillane, and Altreuter, 2006; Spenader, Smits, and Hendriks, 2009).

If Principle A and Principle B bear the same status within the grammar, as was the assumption behind the original binding principles, it remains unexplained why children do not acquire these two principles at approximately the same speed. Also, there is no obvious reason why problems with Principle B should only emerge in interpretation and not in production. For this reason, Hendriks and Spenader (2004; 2005/6) argue that children's asymmetric pattern as well as the adult symmetric pattern of pronominal binding should be explained within a direction-sensitive grammar such as OT (but see e.g. Thornton and Wexler, 1999, and Reinhart, in press, for alternative explanations). In Hendriks and Spenader's OT model, Principle A is taken to be a constraint of the grammar, whereas Principle B is derived as a side effect of bidirectional optimization. Because the two principles have a different status, their different pattern of acquisition can be explained. To see how a direction-sensitive grammar such as OT accounts for both the child pattern and the adult pattern, let us first consider Hendriks and Spenader's analysis of the adult pattern.

3.2 Adult language users and bidirectional optimization

In OT, production is modeled as optimization from an input meaning to the optimal form for expressing this meaning. Comprehension proceeds in the opposite direction and is modeled as optimization from an input form to the optimal meaning assigned to this form. Optimization takes place over a set of hierarchically ordered constraints, which can be divided into markedness constraints and faithfulness constraints. Faithfulness constraints establish a relation between a particular input and a particular output. Markedness constraints, on the other hand, punish particular outputs irrespective of their input. Because markedness constraints are output oriented, and the output differs in production and comprehension, constraints may have different effects in production and comprehension. This may result in different form-meaning pairs in production and comprehension (cf. Smolensky, 1996). However, a symmetric system arises when language users also take into account the opposite perspective. This can be modeled in OT as bidirectional optimization (Blutner, 2000). Bidirectional optimization combines the direction of production with the direction of comprehension and is defined as follows:

(5) Strong bidirectional optimization (adapted from Blutner, 2000):

A form-meaning pair <f,m> is bidirectionally optimal iff:

a. there is no other pair <f',m> such that <f',m> is more harmonic than <f,m>.

b. there is no other pair <f,m'> such that <f,m'> is more harmonic than <f,m>.

Here, bidirectional optimization must be understood as bidirectional evaluation (which will be contrasted with bidirectional learning in section 4). The term 'harmonic' in this definition indicates how well an output candidate satisfies the constraints of the grammar. Under this definition, forms and meanings are not considered separately. Instead, optimization occurs over pairs consisting of forms and their corresponding meanings. A form-meaning pair is an optimal pair if there is no pair with a better form or a better meaning. Only optimal pairs are realized in language. Such optimal pairs block all other pairs in the same competition.

The relative harmony of form-meaning pairs is determined by the constraints of the grammar. To account for the adult pattern of pronominal binding, Hendriks and

Spenader assume Principle A to be a violable constraint punishing reflexives which are not bound locally. In effect, this constraint is violated by any occurrence of a reflexive with a disjoint meaning. In addition to Principle A, which is a faithfulness constraint establishing a relation between a given input and a particular output, a markedness constraint is assumed that results in a preference for reflexives over pronouns: AVOID PRONOUNS. The constraint AVOID PRONOUNS is violated by any occurrence of a pronoun, irrespective of its meaning. This constraint is argued to belong to the constraint hierarchy REFERENTIAL ECONOMY, which reflects the view that expressions with less referential content are preferred over expressions with more referential content (Burzio, 1998; Wilson, 2001). Since cross-linguistically, pronouns tend to be specified for gender and number but reflexives are often unspecified with respect to these features (well-known cases are Dutch and German), reflexives can be said to have less referential content than pronouns. Thus, REFERENTIAL ECONOMY accounts for why forms with less referential content, such as reflexives, are preferred over forms with more referential content, such as pronouns (and over forms with even more referential content, such as full noun phrases).

Given the choice between a reflexive and a pronoun as the form to be selected, and between a conjoint meaning (i.e., a meaning according to which the reflexive or pronoun is coreferential with the local subject) and a disjoint meaning as the meaning to be selected, there are four logically possible form-meaning pairs. These pairs are listed in the first column of bidirectional optimization Tableau 1. Constraints in an OT tableau are ordered from left to right in the first row, in order of descending strength. The linear order of the two constraints indicates that PRINCIPLE A is stronger than AVOID PRONOUNS.

		Principle A	Avoid Pronouns
✌	<reflexive, conjoint>		
	<reflexive, disjoint>	*	
	<pronoun, conjoint>		*
✌	<pronoun, disjoint>		*

Tableau 1: Hendriks and Spenader's bidirectional account of pronominal binding

A crucial property of OT is the violability of constraints. Constraints are potentially conflicting and hence they must be violable. If two constraints are in conflict, it is more important to satisfy the stronger constraint than it is to satisfy the weaker constraint. Because the first pair, <reflexive, conjoint>, satisfies all constraints, whereas all other pairs violate one of the constraints, this first pair is a bidirectionally optimal pair according to the definition given in (5). This is marked by ✌ in the tableau. There is no other pair that satisfies the constraints better (i.e., that is more harmonic). As a result, this first pair blocks all other pairs with the same form but a less harmonic meaning (in this example, the second pair) and pairs with the same meaning but a less harmonic form (in this example, the third pair). Importantly, according to the definition given in (5) also the fourth pair <pronoun, disjoint> is bidirectionally optimal. It does not have any competitors with a more harmonic form or a more harmonic meaning. The third pair is not more harmonic than the fourth pair (instead, it is equally harmonic), and the first pair does not compete with the fourth pair because they have no form or meaning in common. As a result of this bidirectional competition, reflexives are predicted to be used for conjoint meanings and vice versa, and pronouns are predicted to be used for disjoint meanings and vice versa.

Bidirectional optimization Tableau 1 provides a representation of the interpretation of pronouns and reflexives under the assumption that a hearer takes into account the speaker's choices. When a hearer encounters a pronoun, the conjoint interpretation is ruled out because the speaker would have produced a reflexive (which is the optimal form from the speaker's perspective) if she had intended to bring across a conjoint meaning. An alternative characterization of bidirectional optimimization, instead of the definition in (5), is as a sequential process consisting of two steps of unidirectional optimization (cf. Van Rij, Van Rijn and Hendriks, 2010): first, the hearer optimizes from an input form to its optimal meaning (f → m). In a second step, the hearer takes into account the speaker by optimizing in the opposite direction, thereby using the output meaning of the first step as the input to the second step (f → m → f). Whereas adults optimize bidirectionally, Hendriks and Spenader argue that children are unable to do so (see also De Hoop and Krämer, 2005/6) and only apply the first step in interpretation. As a result, children's optimal interpretation may be different from adults' optimal interpretation. Also, children's production may not always yield the same form-meaning pair as their comprehension. These two points will be illustrated in the next subsection.

3.3 Child language users and unidirectional optimization

According to Hendriks and Spenader (2004; 2005/6), children fail to optimize bidirectionally and only optimize unidirectionally. That is, given a particular input, children only consider potential outputs for this input, and do not consider alternative inputs. As can be seen from the tableaux below, children's correct pattern of production of pronouns and reflexives as well as children's errors in the comprehension of pronouns are predicted by Hendriks and Spenader's OT model. The unidirectional Tableaux 2-5 are derived from bidirectional Tableau 1 by considering only pairs with the given input, while everything else is kept constant. For example, Tableau 2 is derived from Tableau 1 by only considering the first and the third row of Tableau 1, in which a particular input meaning (in this case, the conjoint meaning) is represented. The second and fourth row of Tableau 1, which represent an alternative to the input meaning (namely the disjoint meaning), are not relevant for a unidirectionally optimizing language user wishing to express a conjoint meaning. Thus the only difference between Tableau 1, on the one hand, and Tableaux 2-5, on the other, is the mode of optimization. The grammar, i.e., the constraints and their ranking, is the same.

If the input is a conjoint meaning (Tableau 2), the optimal form for expressing this meaning is a reflexive because choosing a pronoun would violate AVOID PRONOUNS. Unidirectionally optimal candidates are marked by ☞.

	Input: conjoint meaning	PRINCIPLE A	AVOID PRONOUNS
☞	<reflexive, conjoint>		
	<pronoun, conjoint>		*!

Tableau 2: Unidirectional production of conjoint meanings

If the input is a disjoint meaning (Tableau 3), on the other hand, only the second and fourth row of Tableau 1 are relevant. In this situation, the optimal form is a pronoun. Even though choosing a pronoun would violate AVOID PRONOUNS, choosing a reflexive would violate the stronger constraint PRINCIPLE A. Hence, a pronoun is the optimal form for expressing a disjoint meaning.

	Input: disjoint meaning	PRINCIPLE A	AVOID PRONOUNS
	<reflexive, disjoint>	*!	
☞	<pronoun, disjoint>		*

Tableau 3: Unidirectional production of disjoint meanings

In comprehension, the input is a form and the output is the optimal interpretation of this form. If the input is a reflexive (Tableau 4), the output is a conjoint meaning. Choosing a disjoint meaning would result in a violation of PRINCIPLE A, whereas choosing a conjoint meaning satisfies both constraints.

	Input: reflexive	PRINCIPLE A	AVOID PRONOUNS
☞	<reflexive, conjoint>		
	<reflexive, disjoint>	*!	

Tableau 4: Unidirectional interpretation of reflexives

The three unidirectional tableaux above yield the same results as bidirectional Tableau 1, predicting adult-like performance for children with respect to the comprehension of reflexives and the production of reflexives and pronouns. However, not in all cases are children predicted to perform adult-like. If the input is a pronoun (Tableau 5), the output of unidirectional optimization is different from the pattern produced under bidirectional optimization.

	Input: pronoun	PRINCIPLE A	AVOID PRONOUNS
☞	<pronoun, conjoint>		*
☞	<pronoun, disjoint>		*

Tableau 5: Unidirectional interpretation of pronouns

Under unidirectional optimization, a pronoun is ambiguous and can be interpreted as expressing a conjoint meaning (the first candidate) as well as a disjoint meaning (the second candidate). These two candidates satisfy and violate the same constraints. Hence they are both optimal. This contrasts with the bidirectional pattern, according to which pronouns are only used for expressing a disjoint meaning and vice versa. Assuming that children cannot represent two different interpretations at the same time and randomly select one of the two optimal interpretations, the difference between unidirectional and bidirectional optimization explains children's guessing pattern when they have to interpret a pronoun. At the same time, children's production of pronouns is predicted to be adult-like (see Tableau 3).

In this section, we discussed a bi-OT model accounting for the acquisition of Principle B. If PRINCIPLE A and REFERENTIAL ECONOMY (AVOID PRONOUNS) are adopted and if it is assumed that adults optimize bidirectionally whereas children cannot do so yet, this bi-OT model is able to explain the effects of Principle B in adults, as well as children's failure with respect to Principle B in comprehension but not in production. However, a drawback of the model is that it does not explain how the two hypothesized constraints came to be part of the grammar. Also, its explanatory power is limited to languages such as English and Dutch. However, see Hendriks, Siekman, Smits and Spenader (2007) for a possible extension of the model that accounts for well-known exceptions to the complementary distribution of pronouns and reflexives as well as for the behavior of pronouns and reflexives in some other languages, by reformulating PRINCIPLE A as a constraint hierarchy of constraints with binding domains of different sizes. For the sake of clarity, in this paper we focus on English. In the next section, we turn to language change and discuss Mattausch's bi-OT model, which presents an account of the emergence of Principle B as a constraint of the grammar of English.

4. A bi-OT model of language change

Language evolution and language change are notoriously difficult to investigate. Relevant data with respect to older stages of modern languages are scarce, and even if it is possible to give an account of *what* has changed in the language, it seems almost impossible to determine *why* it changed. Fortunately, computational models can help investigating the causes of language change. In such models it can be studied whether certain fundamental assumptions give rise to the observed pattern of language change or not.

One such computational model of language change is Mattausch's (2004) bi-OT model of the development of pronominal binding. Mattausch assumes that statistical frequencies in the language play a crucial role in language change. When there is a statistical asymmetry between two forms or two meanings, with one form or meaning being more frequent than the other, the grammar will change in such a way that the constraints of the grammar militate against the infrequent forms or meanings. In this way changes in the grammar reflect the statistical frequencies in the input.

These effects are obtained by a combination of a particular learning algorithm, a particular configuration of input frequencies, and a bi-OT grammar. These three components of Mattausch's simulation model are discussed in the three subsections below.

4.1 The BiGLA learning algorithm

Mattausch assumes the grammar of a language to consist of constraints of varying strength. Learning a grammar equals learning the strength of the constraints of the grammar. To be able to adapt the strength of the constraints, Mattausch employs the BiGLA learning algorithm (Jäger, 2004). This learning algorithm is based on Boersma's (1998) stochastic OT and his Gradual Learning Algorithm (GLA), and adds to the possibility of bidirectional evaluation already present in the GLA the possibility of bidirectional learning. The BiGLA learning algorithm draws an observation (a form-meaning pair) from a corpus at random. The more frequent the form-meaning pair, the higher the chance it is selected. Taking the meaning of the observed form-meaning pair as the input, the optimal form is determined on the basis of the ranked constraints. The mode of evaluation is asymmetrically bidirectional, which means that speakers take into account hearers but not vice versa (we will return to this type of evaluation in section 5.2). If the optimal form in production is different from the observed form, learning takes place. Constraints preferring the optimal form to the observed form (i.e., constraints promoting an incorrect hypothesis) are then decreased in strength, whereas constraints preferring the observed form to the optimal form (i.e., constraints promoting the correct hypothesis) are increased in strength. Crucially, the same procedure is applied in comprehension, taking the form of the observed form-meaning pair as the input, thus determining the optimal meaning on the basis of the grammar and comparing this optimal meaning to the observed meaning. Consequently, learning according to the BiGLA learning algorithm proceeds in a bidirectional fashion and is simultaneously speaker-oriented and hearer-oriented.

Using the BiGLA learning algorithm, every generation of language learning in Mattausch's simulations consists of 60,000 of such observations. On the basis of these observations, the constraint ranking is adjusted until a stable state is reached where no constraint reranking takes place anymore. After 60,000 observations, the first-generation learner will produce a new corpus of form-meaning pairs, which reflects his new constraint ranking. As a consequence of the new constraint ranking, this new corpus may differ slightly from the initial corpus. The new corpus forms the input for the second-generation learner, who will adjust his constraint ranking on the basis of this new corpus. The corpus produced by the second-generation learner will again form the input for the third-generation learner. This process of iterated learning (cf. Kirby and Hurford, 1997) can thus be applied for several generations. The changing strengths of constraints across generations may reflect the process of language evolution, with certain constraints becoming stronger over time and other constraints becoming weaker over time. Often, but not always, after a number of generations the grammar stabilizes into a certain pattern.

4.2 A hypothetical corpus of Old English

Because observations of form-meaning pairs are drawn from a corpus at random, the statistical properties of the corpus are important for the direction of learning, and hence for the direction of language change. The starting point of Mattausch's computational simulation of language evolution are corpus data taken from Keenan (2001). Keenan found in his survey of Old English sources between 750-1154 that 18% of the locally conjoint object pronouns are *self*-marked. The rest are bare pronouns. To complete the pattern of frequencies of forms and meanings, Mattausch adds two assumptions: First, the ratio of disjoint meanings versus conjoint meanings is 49:1. Second, 18% of the locally disjoint object pronouns are also *self*-marked, for example for reasons of contrast or emphasis. This results in the pattern of frequencies shown in the table in (6).

(6) Frequencies of Keenan's Old English

	pro	pro + *self*
conjoint	1.64 %	0.36 %
disjoint	80.36 %	17.64 %

These frequencies form the input for the computational simulations Mattausch presents in his dissertation. Note that this pattern shows no correlation between *self*-marking and whether the meaning expressed is conjoint or disjoint. Conjoint and disjoint meanings are associated with a *self*-marked form equally often (namely in 18% of all cases).

4.3 An Optimality Theoretic grammar

A third and final aspect of Mattausch's computational simulation of the evolution of pronominal binding is the built-in grammar. Mattausch's bi-OT grammar consists of four so-called bias constraints and one markedness constraint. The four bias constraints resemble faithfulness constraints in that they relate a given input to a particular output, or rather prohibit the association between a given input and a particular output:

(7) **self,co*: Self-marked pronouns are not locally conjoint. (= anti-Principle A)
**self,dis*: Self-marked pronouns are not locally disjoint. (= Principle A)
**pro,co*: Bare pronouns are not locally conjoint. (= Principle B)
**pro,dis*: Bare pronouns are not locally disjoint. (= anti-Principle B)

These bias constraints form a "comprehensive pool of codistributional constraints" (Mattausch, 2004: p. 105), referring to all possible form-meaning combinations. In addition to these four bias constraints, Mattausch also adopts a single markedness constraint, *STRUCT, which prohibits the use of morphological structure. This constraint represents speaker economy, and is violated by any *self*-marked form. Note that the effect of *STRUCT is the opposite of the effect of Hendriks and Spenader's constraint REFERENTIAL ECONOMY, which prefers *self*-marked forms to bare pronouns. Because reflexives violate *STRUCT whereas pronouns do not, Mattausch takes reflexives to be the marked forms.

So Mattausch's computational simulation uses the asymmetric BiGLA learning algorithm, corpus frequencies approximating Old English, and a bi-OT grammar consisting of four bias constraints and one markedness constraint. If the simulation is run for several generations, where the output of one generation forms the input for the next generation, the result is a clear ordering of the constraints:

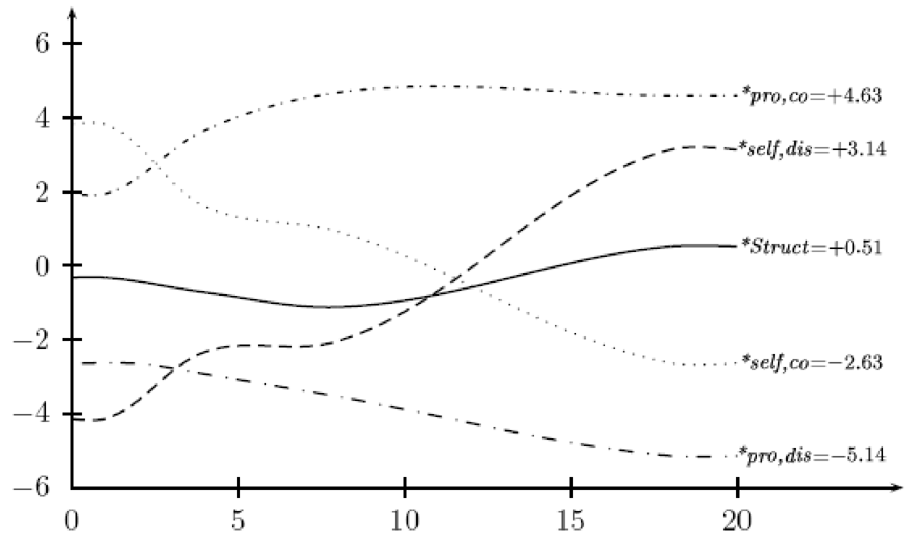

Figure 1: Mattausch's computational simulation of the evolution of pronominal binding, with generations 1-20 plotted along the x-axis and the ranking values of the constraints along the y-axis.[2]

As Figure 1 shows, after 20 generations of learning, the constraint **pro,co* (PRINCIPLE B) emerges as the strongest of the five constraints. The constraint **self,dis* (PRINCIPLE A) has also become a very strong constraint. In other words, the initial tendencies represented by these constraints have been grammaticalized, or 'fossilized'. The constraint *STRUCT is in the middle, and the other two constraints (ANTI-PRINCIPLE A and ANTI-PRINCIPLE B) are ranked lowest. The corpus produced after 20 generations reflects the rule-like behavior of PRINCIPLE B and PRINCIPLE A:

(8) Frequencies of Mattausch's model after 20 generations

	pro	pro + *self*
conjoint	0 %	2 %
disjoint	98 %	0 %

[2] Figure 1 is taken from Mattausch (2004: Fig. 6.40, p. 130). As one of the reviewers pointed out, the curves look continuous, which they should not if they are the result of iterated learning with one teacher and one learner at the time. We suspect that the results have been averaged over a number of subsequent observations.

Reflexives only occur with conjoint meanings, and pronouns only occur with disjoint meanings. Mattausch's simulation results thus show that, under particular assumptions with respect to learning, input frequencies and grammar, PRINCIPLE A and PRINCIPLE B emerge as strong constraints.

Mattausch thus provides an account of the mechanism of grammaticalization that Levinson (2000) claims to be responsible for the three diachronic stages through which languages gradually develop reflexives. In Levinson's stage 1, the language has no reflexives, and bare pronouns are used reflexively. Stage 2 shows the gradual emergence of reflexives (based on e.g. body-part expressions or emphatics), which however coexist with the reflexive use of pronouns. In stage 3, finally, bare pronouns are not used reflexively anymore. Whereas Old English is an example of a stage 1 language, Modern Standard English is an example of a stage 3 language.

5. Comparing the two models

Although the model proposed by Hendriks and Spenader (2004; 2005/6) and the model proposed by Mattausch (2004) are intended to explain different sets of observations with respect to pronominal binding, a comparison of the two models reveals a number of similarities. Both models are formulated within the framework of bidirectional Optimality Theory and proceed from the assumption that language interpretation and generation are the result of optimization over a hierarchically ordered set of constraints. Moreover, they both assume speakers to take into account hearers when producing a linguistic form. Also, both models yield an explanation for why reflexives receive a conjoint interpretation (Principle A), whereas pronouns receive a disjoint interpretation (Principle B). These similarities suggest that the two models may be compatible and can perhaps be combined into a single model which is able to explain the emergence of the binding principles within a single individual as well as within the language itself.

However, a comparison between the two models also reveals a number of differences, of which the following two are perhaps the most important: (i) the nature and choice of the constraints, and (ii) the type of optimization employed. To determine whether the two models can be combined into a single model, we will take a closer look at these differences and see whether these differences can be resolved while maintaining the explanatory power of each model.

5.1 Constraints on pronominal binding

A first difference between Hendriks and Spenader's bi-OT account and Mattausch's bi-OT account concerns the constraints employed. Whereas Hendriks and Spenader's model employs one faithfulness constraint (PRINCIPLE A) and one markedness constraint (AVOID PRONOUNS), Mattausch's model uses four bias constraints (PRINCIPLE A, PRINCIPLE B, ANTI-PRINCIPLE A and ANTI-PRINCIPLE B) and one markedness constraint (*STRUCT). Would it be possible for one set of constraints to explain both the acquisition data and the evolution data? In other words, could the acquisition data be explained by Mattausch's constraints, or the evolution data be explained by Hendriks and Spenader's constraints?

If we employ Hendriks and Spenader's non-frequentist symmetric bidirectional model, but replace their stipulated constraints by Mattausch's bias constraints and his *STRUCT under the ranking derived in Figure 1, we would indeed predict the correct adult pattern:

		PR. B	PR. A	*STRUCT	ANTI-PR. A	ANTI-PR. B
✌	<reflexive, conjoint>			*	*	
	<reflexive, disjoint>		*	*		
	<pronoun, conjoint>	*				
✌	<pronoun, disjoint>					*

Tableau 6: Bidirectional optimization with Mattausch's constraints

However, looking at children's unidirectional interpretation of pronouns under the same constraint ranking, we now incorrectly predict that for children, pronouns must be disjoint to the local subject, too:

	Input: pronoun	PR. B	PR. A	*STRUCT	ANTI-PR. A	ANTI-PR. B
	<pronoun, conjoint>	*!				
☞	<pronoun, disjoint>					*

Tableau 7: Unidirectional interpretation of pronouns with Mattausch's constraints

In fact, we would predict the same pattern for children and adults in all cases, which raises the question whether bidirectional optimization is required for interpretation at all. Indeed, Mattausch's model assumes hearers to optimize unidirectionally, as is a property of the asymmetric BiGLA algorithm and Jäger's EvolOT implementation of this algorithm (see section 5.2).

Perhaps we can explain children's deviant pattern of pronoun interpretation from Mattausch's constraints by assuming that both children and adults optimize unidirectionally but children haven't acquired the adult constraint ranking yet. That is, perhaps the adult pattern is given by Tableau 7 but children's pattern results from a different ranking of the same constraints. That this is possible is shown by the following tableau:

	Input: pronoun	PR. A	PR. B	ANTI-PR. B	*STRUCT	ANTI-PR. A
☞	<pronoun, conjoint>		*			
☞	<pronoun, disjoint>			*		

Tableau 8: Unidirectional interpretation of pronouns under a non-adult ranking of Mattausch's constraints

So it may be that for children, PRINCIPLE A and ANTI-PRINCIPLE B are still ranked too high. ANTI-PRINCIPLE B and PRINCIPLE B must be tied to allow for both interpretations. PRINCIPLE A must be ranked above ANTI-PRINCIPLE B to prevent disjoint meanings to be expressed by a reflexive. To acquire the adult pattern of forms and meanings, then, children must rerank their constraints on the basis of the received input. However, this raises two important questions: (i) why does it take children so long to arrive at the adult constraint ranking?, and (ii) where does the constraint ranking in Tableau 8 come from?

Let's first consider the question why it takes children so long to arrive at the adult constraint ranking. Adult-like comprehension of pronouns generally does not emerge until after the age of 5, which is extremely late for first language acquisition. This is surprising because relevant observations on the basis of which ANTI-PRINCIPLE B can be demoted (viz., pronouns with a disjoint meaning) are by no means rare. This late delay in comprehension, in combination with the different time course of production and comprehension, with correct comprehension of a linguistic item sometimes lagging behind its correct production several years, has motivated two-stage models of language acquisition (De Hoop and Krämer, 2005/6; Hendriks and Spenader, 2004; 2005/6). According to these models, children start out with unidirectional optimization, which causes the well-known acquisition delay with pronouns. Only later does the ability to apply bidirectional optimization to pronouns emerge, perhaps as a result of increased working memory capacity, sufficient speed of processing (see Van Rij, Van Rijn and Hendriks, 2010, for evidence in this direction), or the development of the ability to apply Theory of Mind reasoning to pronominal utterances. This is expected to take time. But if the acquisition of pronominal binding merely is a matter of reranking two constraints, it is not expected that this should take several years.

A second question arising from an explanation of the Delay of Principle B Effect in terms of constraint reranking is where the constraint ranking in Tableau 8 comes from. Apparently, the ranking in Tableau 8 is the ranking that explains children's pattern best. However, this ranking is not the ranking that matches the frequencies in the language best, since the adult ranking is already assumed to do so. So, given that the constraint ranking tries to reflect the statistical properties of the language, where does children's ranking come from? To shed more light on this issue, let us look at Mattausch's simulation of the acquisition of Modern English. The last generation for which Mattausch explicitly presents the learning curves is a thirteenth-generation learner (Mattausch, 2004; Fig. 6.35, p. 128). The final constraint ranking for this generation is not yet the ranking giving rise to Modern English, however. According to the ranking of a thirteenth-generation learner, namely PRINCIPLE B >>

ANTI-PRINCIPLE A >> *STRUCT >> PRINCIPLE A >> ANTI-PRINCIPLE B, reflexives receive a disjoint interpretation. Thus at this stage in language evolution, Principle B is almost fully instated in the grammar, whereas Principle A is not yet present. Since we are interested in the pattern of acquisition of Modern English, in which both Principle B and Principle A are fully instated in the adult language, we ran our own simulation.

Figure 2 shows the results of our simulation of the learning curves of a single learner exposed to a corpus of Modern English.

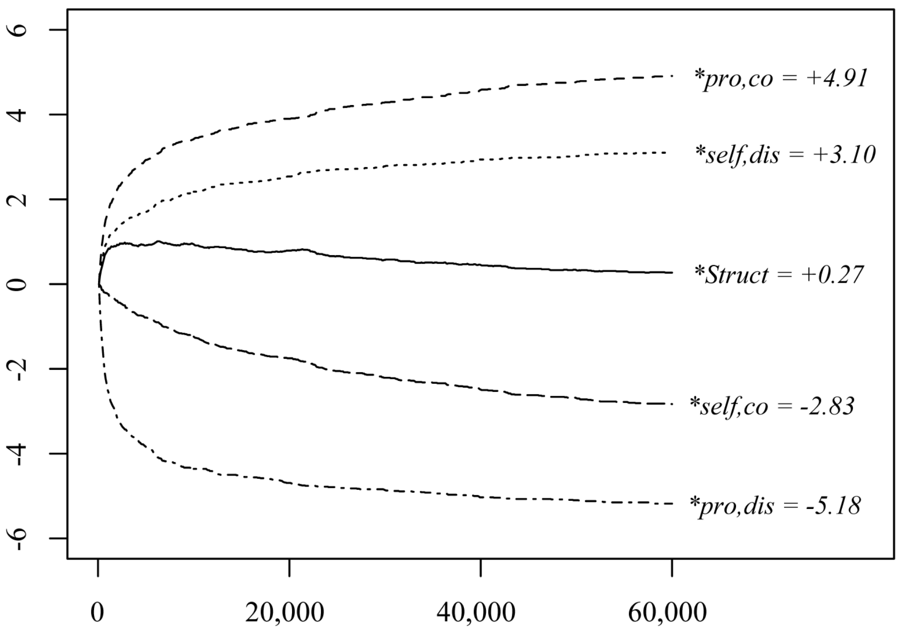

Figure 2: Our bidirectional learning curves of a twentieth-generation learner. Parameter values used: 60,000 observations, a step size of 100 observations, bidirectional mode of evaluation, bidirectional mode of learning, noise value of 2.0 (default), and plasticity value of 0.01 (default).[3]

Figure 2 differs from Figure 1 in that it shows the changes in the strength of the constraints for one individual learner across observations (i.e., during the course of

[3] The curves in Figure 2 and Figure 3 below are less smooth than the curves in Figure 1, because in our simulations we did not average over a number of subsequent observations.

language acquisition), rather than the changes across subsequent generations of learners. The input to the model is a hypothetical corpus of Modern English, as presented in (8). At this stage both Principle B and Principle A are fully grammaticalized in the language.

Each learner starts out with all constraints having the initial value of zero. As can be seen in Figure 2, **pro,co* (PRINCIPLE B) immediately becomes the strongest of the constraints. The learning curves do not show any signs of constraint reranking. The language learner immediately zooms in on a particular constraint ranking, and this ranking does not change anymore over time. Only the strengths of the constraints change. We do not see any evidence of the constraints **pro,co* (PRINCIPLE B) and **pro,dis* (ANTI-PRINCIPLE B) being tied. Rather, PRINCIPLE B and ANTI-PRINCIPLE B immediately diverge strongly. Also, nowhere in the learning curves is **self,dis* (PRINCIPLE A) the strongest of the five constraints, in contrast to what Tableau 8 predicts. Instead, the constraints regulating the behavior of reflexives distance themselves from each other more slowly than the constraints regulating the behavior of pronouns. This means that Mattausch's bi-OT model incorrectly predicts a Delay of Principle A Effect, rather than a Delay of Principle B Effect. So the learning curves resulting from a simulation by Mattausch's model predict the exact opposite of children's actual pattern of acquisition of pronouns and reflexives.

Summarizing, even though Mattausch's bi-OT model is able to explain the adult pattern of pronominal binding, his explanation cannot be extended to account for children's acquisition of pronominal binding. In particular, no Delay of Principle B is predicted, contrary to the facts. Now let us turn to the opposite question and see whether it is possible to explain the hypothesized stages of language change using only Hendriks and Spenader's two constraints PRINCIPLE A and AVOID PRONOUNS. Figure 3 presents our simulation based on only these two constraints.

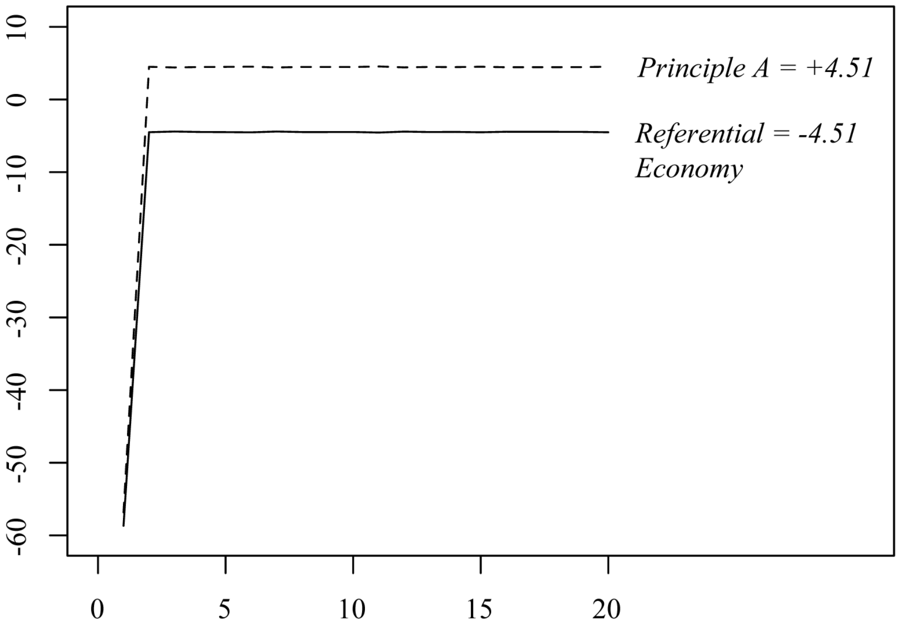

Figure 3: Simulation of the evolution of pronominal binding with the constraints PRINCIPLE A and REFERENTIAL ECONOMY (AVOID PRONOUNS), based on the frequencies of Keenan's Old English. Parameter values used: 60,000 observations, 20 generations, bidirectional mode of evaluation, bidirectional mode of learning, noise value of 2.0 (default), and plasticity value of 0.01 (default).

PRINCIPLE A and REFERENTIAL ECONOMY (AVOID PRONOUNS) were both given an initial value of 0. Figure 3 shows a strong negative value for both constraints at the start of the simulation. This is the result of a strong initial mismatch between the constraints and the corpus. Because of this mismatch, the ranking values of the constraints are adjusted, but as a consequence of the adjustment of the ranking values the output corpus also changes. This then results in a much better fit between the constraints and the corpus. Almost immediately, a stable state is reached where PRINCIPLE A has a positive value and AVOID PRONOUNS is much weaker. Although this ranking corresponds to the constraint ranking assumed by Hendriks and Spenader in their model, the curves do not present a plausible picture of the evolution of pronominal binding. As a result of the ranking values of the constraints as depicted in Figure 3, the output corpus changes from Old English to Modern English in only one generation.

The inadequacy of Hendriks and Spenader's constraints to explain the hypothesized stages of language change is not surprising, because in their model the constraints do only half of the work. Recall that Hendriks and Spenader assume that the adult pattern of forms and meanings observed for pronouns and reflexives in Modern English not only requires PRINCIPLE A to outrank AVOID PRONOUNS, but also requires hearers to take into account the speaker, which is modeled as symmetric bidirectional optimization. As a consequence of this mechanism of bidirectional optimization, a hearer is able to block the non-adult conjoint interpretation of pronouns. This part of linguistic knowledge is not embodied in the constraints of the grammar. Neither does it result from the asymmetric bidirectional evaluation procedure as implemented in the EvolOT software, as hearers do not take into account speakers in this asymmetric evaluation procedure. Therefore, no Principle B effects are expected to arise. Because the evaluation procedure plays such a crucial role, we should not consider the constraints separately but rather consider them in combination with the proposed mechanism of optimization. In the next section, we therefore discuss different options for evaluating candidate outputs.

5.2 Types of optimization

The previous section focused on the different sets of constraints Hendriks and Spenader and Mattausch employ to account for the adult pattern of pronominal binding. A second difference between Hendriks and Spenader's bi-OT account and Mattausch's bi-OT account is the type of optimization. Under Hendriks and Spenader's account, it is crucial that hearers take into account speakers. Recall that Hendriks and Spenader start out with Principle A only. Principle B effects are derived through bidirectional optimization. In their model, pronouns are ambiguous between a conjoint and a disjoint meaning. To be able to interpret a pronoun correctly, hearers have to consider the perspective of a speaker. Because the speaker could have used a reflexive but did not, the hearer may (implicitly) conclude that the speaker did not want to express a conjoint meaning. As a result, the pronoun receives a disjoint interpretation, which corresponds to the effects of Principle B.

Mattausch, in contrast, adopts the asymmetric bidirectional model of BiGLA (Jäger, 2004) which is implemented in the EvolOT software. According to the BiGLA algorithm, speakers take into account hearers in the sense that produced forms must in principle be recoverable, but hearers do not similarly take into account speakers. The choice for such an asymmetric version of bidirectional OT is motivated by Mattausch (2004: fn. 13, p. 90) by pointing out that this allows one to avoid certain puzzles faced by the symmetric version, such as the Rat/Rad problem. This problem arises from the fact that in German, both *Rat* 'council' and *Rad* 'wheel' are pronounced as [rat]. Because the underlying form /rat/ is more faithful to the surface form [rat] than is the underlying form /rad/, under a symmetric version of bidirectional OT it is incorrectly predicted that /rad/ is blocked as a potential underlying form. However, a solution for the Rat/Rad problem is also available within a symmetric version of bidirectional OT (as Bouma, 2008, shows in the domain of syntax). Another possibility that would solve the problem would be to assume more than two levels of representation (see Boersma, 2001, for this type of solution in the domain of phonology). Although an asymmetric bidirectional model does not suffer from the Rat/Rad problem, a drawback of an asymmetrical bidirectional model is that it cannot straightforwardly account for the pattern of marked-forms-for-marked-meanings through partial blocking. Only under particular assumptions regarding the corpus frequencies and the constraints employed does this pattern of marked-forms-for-marked-meanings emerge as the outcome of the evolutionary process, as Mattausch shows for pronouns and reflexives.

In the EvolOT implementation of the BiGLA algorithm, various OT models can be simulated by changing the parameters accordingly. A first parameter concerns the choice between unidirectional and bidirectional evaluation. Under unidirectional evaluation, a form or meaning is evaluated without taking into account alternative forms and meanings. Bidirectional evaluation takes into account alternative forms and meanings, albeit in an asymmetrical fashion. A second parameter concerns the choice between unidirectional and bidirectional learning. Under unidirectional learning, the model learns in the production mode only (i.e., as a speaker) or in the comprehension mode only (i.e., as a hearer). Under bidirectional learning, the model learns both in the production mode and in the comprehension mode. By changing the parameters, evaluation can proceed from meaning to form ($m \rightarrow f$), from form to meaning ($f \rightarrow m$), or both. Consequently, not only the BiGLA model Mattausch employs, but also several other OT models can be implemented and simulated.

Parameters in EvolOT:		**Type of optimization:**		**OT model:**
Evaluation	**Learning**	**Speaker**	**Hearer**	
unidirectional	unidirectional (hearer mode)	m → f	f → m	Boersma & Hamann (2008)
unidirectional	bidirectional	m → f	f → m	Children, according to Hendriks & Spenader (2005/6)
asymmetric bidirectional	bidirectional	m → f → m	f → m	Boersma (1998)[4]; Jäger (2004); Mattausch (2004)
n.a.	n.a.	m → f → m	f → m → f	Adults, according to Hendriks & Spenader (2005/6)
n.a.	n.a.	m → f → m	f ← m	Mattausch & Gülzow (2007)

Table 1: Various OT models differing in the type of optimization, with the corresponding parameter settings in EvolOT for evaluation and learning. The models in the bottom two rows cannot be represented in the version of EvolOT used in the present study.

In the version of EvolOT used in the present study, it is not possible to simulate the symmetric bi-OT model proposed by Hendriks and Spenader or the revised asymmetric bi-OT model proposed by Mattausch and Gülzow (2007) that will be discussed below. As a consequence, we could not determine straightforwardly

[4] In the classification in Table 1, we abstract away from the fact that Boersma (1998) utilizes three instead of two levels of representation. Because of the three levels of representation in his models, EvolOT cannot actually model Boersma's learning algorithms. Also, learning is not straightforwardly bidirectional in Jäger's sense.

whether Hendriks and Spenader's symmetric bi-OT model is able to account for the diachronic development of Principle B. However, using EvolOT we could determine the effects of the inverse situation, where Mattausch's asymmetric bi-OT model is applied to language acquisition. Whether we combine the asymmetric bi-OT model with Mattausch's bias constraints and *Struct (Figure 2) or with Hendriks and Spenader's two constraints Principle A and Referential Economy (Figure 3), we do not see any reflection of the Delay of Principle B Effect. In fact, as we already pointed out above, Mattausch's asymmetric bi-OT model employing bias constraints predicts exactly the opposite of a Delay of Principle B Effect, namely a Delay of Principle A Effect.

5.3 Reformulating bidirectional optimality

Mattausch and Gülzow (2007) acknowledge that the Delay of Principle B Effect presents a serious challenge to the evolutionary bi-OT account proposed in Mattausch (2004). To solve the problem posed by the Delay of Principle B Effect, they reformulate the definition of bidirectional optimality and propose that hearers do not interpret expressions according to interpretational constraints, but according to what a speaker would do if he wanted to express a certain meaning.

(9) Mattausch and Gülzow's revised bidirectional optimality (2007: p. 349):

- a. A meaning m is *recoverable* from a form f iff there is no form-meaning pair <f,m'> such that <f,m'> is more harmonic than <f,m>.
- b. A form-meaning pair <f,m> is *speaker optimal* iff either
  - (i) m is recoverable from f and there is no pair <f',m> such that m is recoverable from f' and <f',m> is more harmonic than <f,m>, or
  - (ii) no form x is such that m is recoverable from x and there is no pair <f'm> such that <f',m> is more harmonic than <f,m>.
- c. A form-meaning pair <f,m> is *hearer optimal* iff there is no pair <f',m> such that <f',m> is more harmonic than <f,m>.

According to Mattausch and Gülzow's revised version of asymmetric bidirectional optimality, interpretation is unidirectional but is guided by production constraints only. This is defined by (9c), which states that meaning *m* is optimal for form *f* if and only if *f* is the optimal output for *m*. Production, on the other hand, is assumed to be bidirectional, as can be seen from the recoverability restriction in (9b).

Figure 4 presents the learning curves resulting from Mattausch and Gülzow's revised bidirectional optimality for a single learner exposed to a corpus of Modern English.

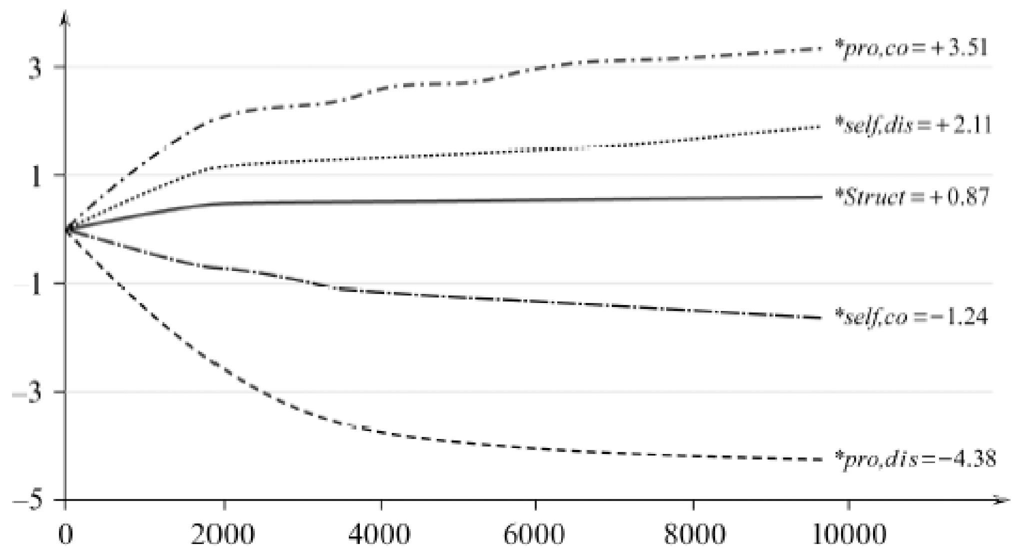

Figure 4: Mattausch and Gülzow's bidirectional learning curves of a learner of Modern English under their revised version of asymmetric bi-OT. The results are after 10,000 inputs. The input frequencies are the ones given in (8).[5]

Because the constraint ranking is continuous, there is a point where the odds that the constraints favor a pronoun for a conjoint interpretation are close to 50%. This point lies at about 5000 learning data, Mattausch and Gülzow mention. This result cannot be read off the learning curves directly but can be determined by running a computational simulation and generating an output corpus, or by calculating the probability of a particular output as a function of the relative ranking values of the constraints determining the output. Because a non-standard relation is assumed

[5] Figure 4 was taken from Mattausch and Gülzow (2007: Fig. 6, p. 350). See footnote 1 for a comment on the smoothness of the curves in this figure.

between production and comprehension, Mattausch and Gülsow's asymmetric bi-OT model cannot be simulated using the current version of EvolOT.

Under Mattausch and Gülzow's account, the Delay of Principle B Effect arises because of stochastic variation. Although PRINCIPLE B (*pro,co*) immediately becomes the strongest of the five constraints, due to stochastic variation there are situations where PRINCIPLE B (*pro,co*) will be outranked by *STRUCT or ANTI-PRINCIPLE A (*self,co*). In the latter case, pronouns will be interpreted as expressing a conjoint meaning. This is expected to happen "somewhere between 25-50% of the time between the ages of, say, four and seven years of age" (Mattausch and Gülzow, 2007: p. 352). The ages mentioned here must probably be seen as a mere indication rather than a concrete prediction, since the strengths of the constraints are a function of the learning data absorbed, not of age. Moreover, the speed of acquisition can be altered by changing the parameters in the computational simulation.

Two important questions arising from Mattausch and Gülsow's account are (i) whether the revised definition of bidirectional optimality they propose yields a plausible model of grammar, and (ii) whether their asymmetric bi-OT model indeed gives an empirically adequate explanation for the phenomenon of pronominal binding. We will discuss the first question immediately below, and the second question in section 5.5.

5.4 The recognition problem

In the previous section, we discussed Mattausch and Gülzow's reformulation of bidirectional optimality, according to which interpretation is guided by production constraints. Although Mattausch and Gülzow do not present independent evidence for this assumption, it allows them to account for children's acquisition of pronouns while maintaining a diachronic perspective. However, their assumption may be problematic, as it seems to suffer from a well-known computational problem which is known as the recognition problem.

As Kuhn (2003, Chapter 6) shows, unidirectional production OT is undecidable with respect to recognition. Recognition for a grammar involves deciding, given an arbitrary string and a grammar, whether the string is part of the language produced by the grammar. In unidirectional production OT, a form is grammatical if that form is an optimal realization of some input. For recognition to succeed, one could start with an arbitrary input meaning and determine whether this input yields the given form as its optimal output. This procedure continues until an input meaning is found that yields the given form as its optimal output. In Mattausch and Gülzow's asymmetric bidirectional model, "when a hearer interprets an expression, he consults his own generative constraints and checks for which meaning that expression is optimal" (p. 350). The procedure they propose thus seems to be similar to the procedure in a recognition task. Since a hearer always starts with the observed form, he must first deduce possible input meanings from this form, then carry out unidirectional production with each of these meanings as the input, and finally check whether one of the optimal forms is identical to the observed form. We can therefore characterize this type of evaluation, which is very similar to recognizing whether a given form is a grammatical form, as the backwards application of unidirectional production: $f \leftarrow m$.

The problem arising with this procedure, which is discussed by Kuhn in relation to the recognition problem, is that we cannot be certain what the relation between a hypothesized input meaning and the given output form is, because faithfulness constraints are violable. As a consequence, any aspect of meaning can remain unexpressed in the output. We cannot use the constraint profile to guide our search, because we do not know how harmonic the form-meaning pair will be. Thus there is no systematic way to explore the infinite space of possible inputs. Consequently, we can never stop looking until we have found an input meaning meeting the requirements, and hence the recognition problem is undecidable for these optimization models.[6] Of course, in Mattausch and Gülzow's computational simulations, where only two meanings are considered, this will not give rise to any problems, but it is problematic for more realistic models of grammar.

Kuhn presents two ways to solve the recognition problem: The first one is to pose restrictions on the input meanings. Obviously, this will restrict the search space of meanings. However, this solution seems to go against the basic idea of frequentist models to assume as few built-in restrictions as possible. The second solution Kuhn presents for the recognition problem is to move to strong bidirectional optimization, that is, the symmetric type of bidirectional optimization employed by Hendriks and

[6] Note that the recognition problem is different from the general OT problem of the infinity of the candidate set (see Kuhn, 2003, for a discussion of the differences, and for a computational solution to the problem of the infinity of the candidate set).

Spenader (2005/6), which is crucially different from the asymmetric type of bidirectional optimization employed by Mattausch and Gülsow. In strong bidirectional optimization, the input meaning that is to be found in the recognition task has to be among the optimal meanings for the given form. This gives us a way of systematically searching the space of meanings (see also Bouma, 2008, for a discussion of this issue).

5.5 Adult processing of pronouns

The bi-OT model of Mattausch (2004) and the revised model of Mattausch and Gülsow (2007) predict that after a certain amount of input data, the language learner will have reached a stable state in which Principle B (**pro,co*) is a strong constraint, which is even stronger than Principle A (**self,dis*). This is illustrated by the learning curves in Figure 2 and Figure 4, respectively. In contrast, Hendriks and Spenader (2004; 2005/6) predict that in adults, Principle A and Principle B are qualitatively different principles of grammar. Whereas Principle A is assumed to be a constraint of the grammar, Principle B is argued to be a derived effect that emerges as the result of bidirectional optimization. As a result, Principle B is expected to be a much more vulnerable cue in interpretation.

Can we find evidence for such a qualitative difference between Principle A and Principle B in adult language in support of Hendriks and Spenader's model and contradicting Mattausch's and Mattausch and Gülsow's models? Indeed, there seems to be evidence for a difference between Principle A and Principle B in language breakdown in aphasia. Grodzinsky, Wexler, Chien, Marakovitz and Solomon (1993) examined anaphoric reference assignment in adult agrammatic Broca's patients (age 58-71) using the same materials as Chien and Wexler (1990) used to test pre-school children. Because the Broca's patients were at least in their late thirties at the time of onset of the aphasia, they can be expected to have fully mastered the grammar of their native language. However, Grodzinsky et al. found that the performance of Broca's patients mirrored the performance of the pre-school children in Chien and Wexler's experiment. Both populations were found to experience fewer problems when interpreting reflexives than when interpreting pronouns in the same constructions. Ruigendijk, Vasić and Avrutin (2006) found a similar pattern with adult Dutch aphasics. So when language breaks down in aphasia, Principle B may be affected while Principle A is still intact, but never vice versa. This suggest that Principle B is of a different nature than Principle A and may require more processing resources than Principle A.

Summarizing, the model proposed by Mattausch (2004) is unable to account for the Delay of Principle B Effect. The revised model of Mattausch and Gülsow (2007) is, but this model has two serious drawbacks: First, as a consequence of the particular relation it assumes between production and comprehension, it suffers from the recognition problem. Second, the model is unable to explain the observation that, even for English adults, Principle B is somehow less 'hard' than Principle A and can be affected while Principle A remains intact.

6. Conclusion

We compared Hendriks and Spenader's (2004; 2005/6) two-stage model of children's acquisition of Principle A and B of Binding Theory with Mattausch's (2004) frequency-based model of the evolution of Principle A and B. Although the two models are both formulated within the framework of bidirectional Optimality Theory, they cannot be combined into a single model of grammar. An explanation of the emergence of Principle B as resulting from statistical patterns in the language seems fundamentally incompatible with the weaker status of Principle B as compared to Principle A in language use, as evidenced by the Delay of Principle B Effect in children and cases of language breakdown in adults. The mismatch between the learning curves of Mattausch's frequency-based model of grammar and children's actual pattern of acquisition suggests that language acquisition may not be solely dependent on statistical patterns in the language. Rather, linguistic knowledge and cognitive properties of the language learner seem to play an important role. On the other hand, linguistic and cognitive factors alone seem unable to explain the process of language change. To arrive at a linguistic theory that is able to account for language acquisition as well as language change, therefore, it seems that both external (statistical) and internal (linguistic/cognitive) factors must be taken into account, albeit in a more sophisticated way than we have considered in this paper.

References

Boersma, P. (2001). Phonology-semantics interaction in OT, and its acquisition. In Kirchner, R., Wikeley, W. and Pater, J., editors, *Papers in Experimental and Theoretical Linguistics*, Vol. 6: 24-35. University of Alberta, Edmonton.

Boersma, P. (1998). *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam.

Boersma, P. and S. Hamann (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology,* 25:217-270.

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

Bouma, G.J. (2008). *Starting a sentence in Dutch: A corpus study of subject- and object-fronting*. PhD dissertation, University of Groningen.

Burzio, L. (1998). Anaphora and Soft Constraints. In: Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M. and Pesetsky, D., editors, *Is the Best Good Enough? Optimality and Competition in Syntax*. MIT Press, Cambridge, MA.

Chien, Y.-C. and Wexler, K. (1990). Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics, *Language Acquisition,* 13:225-295.

Chomsky, N. (1981). *Lectures on government and binding*. Foris, Dordrecht.

De Hoop, H. and Krämer, I. (2005/6). Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition,* 13:103-123.

De Villiers, J., Cahillane, J., and Altreuter, E. (2006). What can production reveal about Principle B? In: Deen, K.U., Nomura, J., Schulz, B. and Schwartz, B.D., editors*The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America*, Honolulu, HI. University of Connecticut Occasional Papers in Linguistics, 4, pp 89-100.

EvolOT, software and user's manual downloadable from Gerhard Jäger's homepage at http://www2.sfs.uni-tuebingen.de/jaeger/evolOT/

Grimshaw, J. and Rosen, S.T. (1990). Knowledge and obedience: The developmental status of the Binding Theory, *Linguistic Inquiry,* 21:187-222.

Grodzinsky, Y., K. Wexler, Y.-C. Chien, S. Marakovitz and J. Solomon (1993). The breakdown of binding relations. *Brain and Language,* 45:396-422.

Hendriks, P., Siekman, I., Smits, E.J. and Spenader, J. (2007). Pronouns in competition: Predicting acquisition delays cross-linguistically. In: Bittner, D. and Gagarina, N., editors, *ZAS Papers in Linguistics*, Volume 48 (Intersentential Pronominal Reference in Child and Adult Language. Proceedings of the Conference on Intersentential Pronominal Reference in Child and Adult Language), pp. 75-101.

Hendriks, P. and Spenader, J. (2004). A bidirectional explanation of the pronoun interpretation problem. In: Schlenker, P. and Keenan, E., editors, *Proceedings of the ESSLLI '04 Workshop on Semantic Approaches to Binding Theory*, Nancy, France.

Hendriks, P. and Spenader, J. (2005/6). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition,* 13:319-348.

Jäger, G. (2004). Learning constraint sub-hierarchies: The bidirectional gradual learning algorithm. In: Blutner, R. and Zeevat, H., editors, *Optimality Theory and Pragmatics* (pp. 251-287). Palgrave/Macmillan, Houndmills, Basingstoke, Hampshire.

Keenan, E. (2001). Explaining the creation of reflexive pronouns in English. Unpublished manuscript, UCLA.

Kirby, S. and Hurford, J. (1997). The evolution of incremental learning: Language, development and critical periods. Technical report, Language Evolution and Computation Research Unit, University of Edinburgh.

Kuhn, J. (2003). *Optimality-Theoretic Syntax - A declarative approach*. CSLI Publications, Stanford, CA.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press, Cambridge, MA.

Lightfoot, D. (1999). *The development of language: Acquisition, change, and evolution*. Blackwell, Oxford, England.

Mattausch, J. (2004). *On the optimization and grammaticalization of anaphora*. PhD dissertation, Humboldt Universität Berlin.

Mattausch, J. and Gülzow, I. (2007). A note on acquisition in frequency-based accounts of Binding Phenomena. In: Gülzow, I. and Gagarina, N., editors, *Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept*. Mouton de Gruyter, p. 331-357.

Pearl, L. and Weinberg, A. (2007). Input filtering in syntactic acquisition: Answers from language change modeling. *Language Learning and Development,* 3:43-72.

Reinhart, T. (in press). Processing or pragmatics? Explaining the coreference Delay. In: Gibson, E. and Pearlmutter, N.J., editors, *The processing and acquisition of reference*, MIT Press, Cambridge, MA.

Ruigendijk, E., Vasić, N., and Avrutin, S. (2006). Reference assignment: Using language breakdown to choose between theoretical approaches. *Brain and Language,* 96:302-317.

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

Spenader, J., Smits, E.J., and Hendriks, P. (2009). Coherent discourse solves the Pronoun Interpretation Problem. *Journal of Child Language,* 36:23-52.

Thornton, R. & Wexler, K. (1999). *Principle B, VP ellipsis and interpretation in child grammar*. MIT Press, Cambridge, MA.

Van Rij, J., Van Rijn, H., and Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language,* 37:731-766.

Wilson, C. (2001). Bidirectional optimization and the theory of anaphora. In: Legendre, G. and Vikner, S., editors, *Optimality-theoretic Syntax*. pp. 465-507. MIT Press, Cambridge, MA.