ACQUISITION AND OPACITY

OLGA TIHONOVA

Master's Thesis in Theoretical Linguistics (LIN-3990)
Department of Language and Linguistics
Faculty of Humanities
University of Tromsø
Spring semester 2009

Table of Contents

Acknowledgements

I would like to use this opportunity to express my profound gratitude to professors at the University of Tromsø, who, over the past two years, generously shared with me their skills and expertise in all areas of phonology.

I am especially indebted to my supervisor, Dr. Martin Krämer. Without his wise guidance and invaluable assistance, this work could never have been written. During my thesis year, Dr. Krämer provided not only fruitful discussions and to-the-point criticism, but also encouragement and reassurance.

Finally, and most importantly, I am deeply grateful to my family for their endless patience, understanding, moral support and all imaginable help that they provided.

Abstract

This thesis focuses on the most recent OT-based theory of opacity called *Optimality Theory with Candidate Chains* (OT-CC, see McCarthy 2007). To date very little attention has been dedicated to the problem of acquisition of OT-CC grammars and to the treatment of spontaneous opacity effects in the light of OT-CC. In this thesis we demonstrate that OT-CC grammars can be effectively learned by the BCD algorithm (Prince & Tesar 2004). Also, on the basis of evidence from obligatorily counterbleeding processes, NDEBs and non-target-like opacity effects, we propose to make certain changes to the status of Precedence constraints with the view to increase the descriptive adequacy of OT-CC. We show that our proposed adjustments allow OT-CC to account for emergence and subsequent loss of spontaneous opacity effects, as well as for the phenomenon of U-shaped learning and cross-subject variation in early production data.

Chapter 1: Introduction

This thesis deals with a phenomenon that causes heated controversy among the proponents of different currents in phonological theorizing; the phenomenon that can shed light on such long-standing problems as language change and acquisition; the phenomenon that has been dubbed "*the* single most important issue in current phonological theory"(Idsardi 2000:337). I am talking, of course, about phonological opacity.

The term 'phonological opacity' refers to cases where a surface form of some language L has exceptionally undergone or failed to undergo a certain phonological process active in language L. The serialist accounts of opacity crucially rely on the existence of some intermediate form that deviates from the input and from the actual output. According to Prince & Smolensky (1993:6), in 'classic' Optimality Theory (OT) the "Input → Output map has no internal structure: all possible variants are produced by Gen in one step and evaluated in parallel". Precisely this property makes opacity a challenging issue for OT.

There have been many attempts to incorporate opacity in OT, which can be roughly subdivided into two large (and partly overlapping) groups: those that rely on expanding some basic assumptions about the nature of underlying representations or the constraint component Con and those that make reference to forms distinct from the input and the actual output. Virtually any of those approaches is associated with a number of problems, which sometimes prove fatal under closer examination. Among those, the acquisition problem is not the least.

Apart from being typologically adequate, any successful theory of grammar must be demonstrated to be learnable by means of an effective computable algorithm. Since OT was proposed in 1993, a number of learning algorithms have been put forward and claimed to solve the problem. The most widely acclaimed and thoroughly tested is, perhaps, the family of algorithms based on *constraint demotion* (CD; see Tesar & Smolensky 1993, Tesar 1995, Prince & Tesar 2004, Boersma 2008). In the course of time, CD algorithms have been shown to successfully account for the acquisition of constraint rankings responsible for transparent generalizations. Moreover, theoretical claims concerning the acquisition progression have been substantiated through empirical research.

Unfortunately, OT-based solutions to the opacity problem are not always compatible with CD algorithms. The matter is further complicated by the fact that opaque generalizations are by no means limited to fully-developed grammars: there is an extensive amount of early production data proving that non-target-like opacity effects spontaneously arise in developing grammars and are lost at the later stages of acquisition. Thus, acquisition and opacity intersect in more than one point: first, one has to account for the acquisition of target-like opacity effects; second, one has to deal with spontaneous emergence and subsequent loss of non-target-like opaque generalizations in the course of acquisition.

The focus of this thesis is the most recent OT-based theory of opacity called *Optimality Theory with Candidate Chains* (OT-CC, see McCarthy 2007), which represents the synthesis of OT with derivations. Although OT-CC has been ambitiously claimed to be "the best theory of opacity - and of phonology generally" (see McCarthy 2007:3), to date very little attention has been dedicated to the problem of acquisition of OT-CC grammars and to the treatment of spontaneous opacity effects in the light of OT-CC. The main goal of this work is to fill the gap. In this thesis we demonstrate that OT-CC grammars can be effectively learned through the BCD algorithm (Prince & Tesar 2004). Moreover, we also show that if certain independently motivated adjustments are made to the theory, OT-CC can successfully deal with spontaneous opacity effects.

This thesis is organized as follows: in Chapter 2 we address some general acquisition problems and provide an overview of currently available solutions; in Chapter 3 we look at a number of OT-based approaches to opacity and their implications for acquisition; in Chapter 4 we focus on OT-CC in its original formulation, while in Chapter 5 we propose certain adjustments to OT-CC with the view to increase its typological adequacy; in Chapter 6 we analyse spontaneous opacity effects in the light of the 'updated' version of OT-CC, while in Chapter 7 we provide an acquisition model of target-like counterbleeding opacity; in Chapter 8 we briefly summarize the discussion.

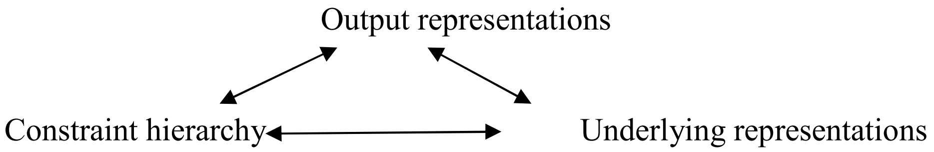
Chapter 2: Learnability in Optimality Theory

2.1 Introduction

Optimality Theory (see Prince & Smolensky 1993) is based on the assumption that UG contains a set of violable constraints, while language-specific grammars are defined through the hierarchical ranking of these constraints. The crucial property of language-specific grammars is their restrictiveness, i.e. when fed some input, such grammar should be able to map it to a surface form that is 'legal' in a particular language. In accordance with the Richness of the Base, the set of inputs to the grammars of all languages is the same. This means that the set of legal outputs for any given language depends solely on the hierarchy of the constraints.

Having acquired the OT grammar, therefore, means having acquired a language-specific ranking of universal violable constraints and correct underlying representations for language-legal outputs. Thus, the task of the language-learner is to pair each surface form with its correct lexical representation and to construct a ranking of the universal constraints that would correctly derive the legal output forms of the target language from any input. The task is further complicated by the fact that all three variables are interdependent. For illustration, consider the scheme in (2-1) below.

(2-1) *Interdependence of three factors in language acquisition* (from Kager 1999a:223)



In other words, in order to pair underlying forms with their output correspondents it is necessary to know the language-specific constraint hierarchy. The constraint hierarchy, in its turn, cannot be constructed unless language-specific input-output mapping in known. To a learned scholar, it might appear to be a vicious circle, while young children seem to cope with the task easily.

The fact that children acquire the language in a relatively short time regardless of all apparent complications and potential traps calls for a formal algorithm whereby we could model the acquisition process. Since OT was proposed as a restrictive and efficient theory of grammar (see Prince & Smolensky 1993), a number of learning algorithms have been devised in order to solve the learnability problem. The most widely acclaimed and thoroughly tested is, perhaps, the family of algorithms based on *constraint demotion* (CD; see Tesar & Smolensky 1993, Tesar 1995, Prince & Tesar 2004, Boersma 2008). In this chapter we will consider currently available CD algorithms, and provide a brief discussion as to their relative merits and drawbacks. The chapter is organized as follows: in Section 2.2 we will address a problem of learning language-specific constraint rankings, abstracting for a while from the need to infer correct lexical forms; in Section 2.3 we will address the problem of learning underlying representations; in Section 2.4 we will briefly summarize our discussion.

2.2 Learning the Constraint Ranking

The first OT-compatible learning algorithm, *Recursive Constraint Demotion* (RCD), was proposed by Tesar & Smolensky 1993. According to Tesar & Smolensky 1993, the ‘raw material’ the language learner starts with consists of a set of universal constraints and a set of well-formed outputs of the target-language (i.e. the adult forms that are observed by the learner). For each optimal output a set of competitors is generated (by means of Gen). Each competitor is *a priori* known to be sub-optimal, hence it is supposed to be ruled out by the constraint ranking of the target grammar. Optimal and sub-optimal forms are ordered pairwise so that each pair contains an optimal output and a competitor. Such pairwise orderings are stored together with the list of violation marks incurred by each form in a pair. Violations incurred by sub-optimal candidates are dubbed ‘loser-marks’, while marks incurred by optimal candidates are dubbed ‘winner-marks’.

The goal of the learner is to find such a stratified hierarchy of constraints that would render each optimal candidate more harmonic than each of its competitors. The algorithm proceeds as follows (where *mark-data* refers to a set of pairs of mark lists):

(2-2) *RCD* (from Tesar & Smolensky 1993:13)

I. Mark Cancellation

For each pair (*loser-marks*, *winner-marks*) in *mark-data*:

a. For each occurence of a mark *C in both *loser-marks* and *winner-marks* in the same pair, remove that occurence of *C from both.

b. If, as a result, no *winner-marks* remain, remove the pair from *mark-data*.

II. Recursive Ranking

a. Output *highest-ranked-constraints* = all the constraints in *not-yet-ranked-constraints* which do not appear in the column *winner-marks* of *mark-data*; these form the highest-ranked stratum of the *not-yet-ranked constraints*.

b. Remove the *highest-ranked-constraints* from the *not-yet-ranked-constraints*.

c. Remove all rows from *mark-data* which contain any marks assessed by the *highestranked-constraints*.

d. Call Recursive Ranking again, with the reduced *mark-data* and the reduced *not-yet-ranked-constraints*.

In a nutshell, according to RCD in the course of acquisition the learner demotes the constraints violated by the intended winner to lower strata to ensure that every violation mark incurred by the intended winner is dominated by the violation incurred by some loser. According to the authors, the RCD is guaranteed to converge upon some stratified grammar that is consistent with all the data encountered by the language learner, provided that such a grammar exists. The result of the algorithm is a stratified hierarchy where the uppermost stratum is occupied by the constraints that are never violated by optimal forms.

However, it was soon observed that, attractive as it looks, the RCD algorithm in its original formulation can lead a learner into a number of traps. For the original version of the Recursive Constraint Demotion algorithm, Tesar and Smolensky (1993:10) assume that at the initial state the constraints are mutually unranked and occupy the only stratum of the child's grammar. According to Smolensky (1996b:7) such a state of affairs proves to be problematic if the learner has to acquire a language L with an unmarked inventory, such that an optimal output form always satisfies both markedness and faithfulness constraints. In the course of acquisition, the learner of language L will only have access to the positive evidence, i.e. the

CD algorithm will be fed optimal outputs only. Since every form of language L satisfies both markedness and faithfulness, no demotions of constraints will be necessary. Eventually, the CD algorithm will converge on a final grammar containing only one stratum of mutually unranked constraints. Any total ranking of these constraints will correctly generate language-specific outputs when provided with language-specific unmarked inputs. However, when presented with a marked input such a grammar will turn out to be too permissive by allowing the marked input to surface faithfully rather than mapping it onto a language-specific unmarked output form.

Another potential problem for RCD is the Identity Map, i.e. the assumption that at the early stages of language acquisition children take underlying representations to be identical with surface representations (see Smith 1973, Gnanadesikan 1995, Smolensky 1996a, Kager 1999a, Prince & Tesar 2004; see also the discussion in Section 2.3 hereof). Hayes (2004:170-175) illustrates on the example of Pseudo-Korean that the Identity Map combined with the RCD outlined above results in a grammar where all faithfulness constraints are top-ranked, because under the Identity Map faithfulness constraints are never violated by the optimal outputs. It is needless to say that such a grammar is far too permissive, allowing any potential input form to surface faithfully.

As such, this result is not necessarily bad. What if the language indeed works this way, and the grammar that we learn is more permissive than it is necessary to account for the data of our mother tongue?

One piece of evidence contradicting this claim comes from a phenomenon known as loanword adaptation. Farris-Trimble (2008:117) provides data from the language Fon, a Gbe language spoken in Benin. According to Farris-Trimble (ibid), the segment inventory of Fon contains only one liquid, namely [l]. No rhotic liquids appear in the language, therefore the constraint Faith[rhotic liquid] is always vacuously satisfied by the native forms. Consistent with our current assumption about the ranking at the initial state, it means that in Fon Faith[rhotic liquid] constraint is high-ranked. What would happen if the language like Fon were to borrow a word containing a rhotic liquid from a language like French or English? The answer is obvious: the high-ranking Faith[rhotic liquid] constraint would require such a loanword to be reproduced faithfully.

However, according to Farris-Trimble, this is far from what happens in reality. It turns out that French and English rhotics are adapted as [l] when they appear in the onset, and deleted altogether when they appear in coda position (due to NoCoda being active in the Fon language).

(2-3) *Onset /ʀ/ is replaced with [l]*

/ʀido/	[lido]	‘curtain’
/byʀo/	[bilo]	‘office’
/gʀɛv/	[glevu]	‘beach’

The empirical study carried out by Smolensky, Davidson & Jusczyk 2004 produced similar results. The study shows that when asked to produce non-English-like clusters English speakers tend to employ different strategies (e.g. schwa-epenthesis) to ‘repair’ such clusters and make them sound ‘English-like’. This suggests that despite the fact that English-language learners were never exposed to such clusters in the process of language acquisition, they are able to detect them as being ‘ill-formed’ in English.

In addition to being empirically untenable, overly permissive grammars generated by RCD also contradict one of the basic tenets of Optimality Theory known as the *Richness of the Base* (Prince & Smolensky 1993:209), whereby the set of inputs to all languages is universal and it is a responsibility of a language-specific grammar to map any possible input fed into it onto an output form that is legal in a particular language.

(2-4) *Richness of the Base* (from Smolensky 1996b:3)

> The source of all systematic cross-linguistic variation is constraint reranking. In particular, the set of *inputs* to the grammars of all languages is the same. The grammatical inventories of a language are the *outputs* which emerge from the grammar when it is fed the universal set of all possible inputs.

According to Smolensky (ibid), the Richness of the Base requires that differences in inventories across the languages result from different constraint rankings, not different inputs. Therefore, given an input that is not a part of the lexicon of language L, a good OT grammar should be able to map it onto the output that would be a legal output in language L. Clearly, the grammars generated by RCD are unable to fulfil this task.

The problem of learning grammars that are too permissive is also known as the Subset Problem (see Prince & Tesar 2004, Smolensky 1996b). According to Prince & Tesar (2004:249), "under OT, the restrictiveness of a grammar depends upon the relative ranking of the constraints", to the effect that "increased domination of markedness constraints over faithfulness constraints will lead to a reduced language consisting of relatively unmarked forms". Therefore, "subset/superset configurations among observable language data can be managed by attention to markedness/faithfulness relationships within the grammar". It was suggested (Smolensky 1996b, van Oostendorp 1995, Sherer 1994) that ranking Markedness over Faithfulness at the Initial State allows to avoid the Subset Problem. Evidence in support of such initial ranking also comes from the empirical study of early production and perception (Gnanadesikan 1995, Smolensky, Davidson & Jusczyk 2004), which shows that children proceed from having extremely restrictive grammars to more permissive ones, gradually expanding their production inventory to include more marked segments.

As it follows from the above discussion, for a grammar to be restrictive it means having faithfulness constraints ranked as low as possible. In order to better capture this desired state of affairs, Prince & Tesar (2004:251) propose to introduce "a numeric metric on constraint hierarchies", which they call *r-measure*.

(2-5) *R-measure* (from Prince & Tesar 2004:252)

> The r-measure for a constraint hierarchy is determined by adding, for each faithfulness constraint in the hierarchy, the number of markedness constraints that dominate that faithfulness constraint.

According to Prince & Tesar (2004: ibid), "any learning algorithm should return a grammar that, among all consistent with the given data, has the largest r-measure". However, it was demonstrated (Hayes 2004, Prince & Tesar 2004) that simply starting out with all Markedness constraints outranking all Faithfulness constraints at the initial state is not enough to achieve this goal. As noted in Prince & Tesar (2004:264), even if at the initial state the ranking is such that $M_1...M_n >> F_1... F_n$, the occurrence of an optimal M_1-violating candidate will be enough evidence for an unbiased learner to establish a hierarchy like $M_2...M_n >> F_1... F_n >> M_1$, reducing the r-measure of the grammar by the total number of universal faithfulness constraints. Such considerations lead Prince & Tesar 2004 (and, independently, Hayes 2004) to conclude that the bias for having Faithfulness constraints low-ranked should hold throughout the language learning.

In order to implement this principle, Prince & Tesar 2004 devised a modified version of the original RCD algorithm, called Biased Constraint Demotion (BCD). The ultimate goal of the algorithm is to prevent the learner from reranking faithfulness constraints unless absolutely necessary, thus ensuring that the resulting grammar has a maximal possible r-measure. The basic modification made to this effect is called 'Faithfulness Delay':

(2-6) *Faithfulness Delay* (Prince & Tesar 2004:259)

> On each pass, among those constraints suitable for membership in the next stratum, if possible place only *markedness constraints*. Only place faithfulness constraints if no markedness constraints are available to be placed in the hierarchy.

At some point in the learning process, the learner might face the situation where the incoming language data cannot be facilitated by ranking of markedness constraints only. The learner then has to choose what constraints from the available faithfulness constraints have to be ranked. According to Prince & Tesar (2004:260), when given such choice the learners should rank "those faithfulness constraints whose ranking will free up markedness constraints for ranking in the next round". Consider Tableau (2-7) below:

(2-7) *Freeing up markedness constraints* (from Prince & Tesar 2004:260)

MDP: Winner ~ Loser	M1	M2	F1	F2
(a) W1 ~ L1	W	L		
(b) W2 ~ L2	L	W	W	

In Tableau (2-7), markedness constraints M1 and M2 cannot be ranked with respect to each other since each of them is violated by some intended winner. Ranking faithfulness constraint F2 will not change this situation, because F2 is unviolated by winners and losers alike. Ranking F1, however, will eliminate the data-pair in (b), thus making it possible to establish the mutual ranking of M1 and M2, such that M1 >> M2. We can then say that by ranking F1 we have *freed up* M1 for further ranking.

Ranking both F1 and F2 is also an option, but it will lead to the unwarranted and undesired decrease of r-measure. In order to maximise the r-measure of the resulting grammar, Prince & Tesar (2004:260) introduce a second ranking principle the learners should follow, called ‘Avoid the Inactive’.

(2-8) *Avoid the Inactive* (from Prince & Tesar 2004:260)

> When placing faithfulness constraints into the hierarchy, if possible only place those that *prefer some winner*. If the only available faithfulness constraints *prefer no remaining winners*, then place all of them into the hierarchy.

Basically, this principle ensures that only those faithfulness constraints that conflict with other constraints are ranked. Prince & Tesar (2004:266) further note, however, that in some cases markedness constraints can be freed up only if two or several faithfulness constraints are ranked together. In order to avoid an unmotivated decrease of r-measure, such cases are handled by means of the following principle:

(2-9) *Smallest Effective F sets* (from Prince & Tesar 2004:267)

> When placing faithfulness constraints into the hierarchy, place *the smallest set* of F constraints that *frees up some markedness constraint*.

If the algorithm is able to find more than one such sets, those that free up more markedness constraints in contiguous subsequent strata are preferred by *Richest Markedness Cascade* principle (see Prince & Tesar 2004:268).

Thus, BCD algorithm is clearly conceptually superior to RCD in that it allows to avoid the Subset Problem and is guaranteed to return a stratified hierarchy with a maximal possible r-measure.

Another departure from the original RCD concerns the method whereby competing sub-optimal forms are obtained. In the original proposal (Tesar & Smolensky 1993:3), sub-optimal competitors are generated by the function Gen. One property of Gen is that the number of sub-optimal forms it creates is infinite. Clearly, among such forms many will violate markedness and faithfulness constraints never violated by optimal candidates. Consequently, the pairwise comparison of such forms with optimal candidates will give no evidence for constraint demotion, i.e. no learning will occur. It follows that the learner might have to go through a potentially infinite list of uninformative winner-loser pairs until she finds evidence to demote some loser-favouring constraint. This point was first made by Tesar (1995:95), who indicates that given the property of Gen to generate an infinite number of suboptimal candidates, "it is not immediately clear how to algorithmically select suboptimal descriptions". Tesar (ibid) proposes Error-Driven Constraint Demotion algorithm (EDCD), which computes the optimal output of the learner's current stratified grammar and selects it as a loser for further pairwise comparison by RCD. To this end, Tesar (1995:96) devises a way whereby the learner can calculate a relative harmony of output candidates given a hierarchy that is only partially ranked.

(2-10) *Extension of Harmonic Ordering of Forms* (from Tesar 1995:96)

> Two descriptions are ordered relative to a single stratum by listing for each description the marks assessed by all the constraints in the stratum. The description with fewer marks is the more Harmonic relative to that stratum. If they have the same number of marks, the two descriptions are not Harmonically distinguished relative to that stratum.

The method of relative harmony assessment whereby the violation marks incurred by the candidate on unranked constraints within a stratum are added up was dubbed 'pooling ties*'*' (by Boersma 2008:4, who attributes the term to Tesar 2000). The loser selected this way is guaranteed to be informative, since, being the output of the learner's grammar at the initial state, it clearly fares better than the intended winner on certain high-ranked constraints. BCD, discussed above, retains this method of loser-selection (see Prince & Tesar 2004:257).

According to Boersma (2008:4), precisely this property leads Error-Driven algorithms to the trap avoided by the earlier RCD variants. Boersma (ibid) considers a hypothetical situation whereby the learner presented with a single language datum computes an output of her current stratified hierarchy that happens to be equivalent to the optimal output. Once it happens, the EDCD algorithm is considered to have converged on a correct stratified hierarchy, with the property that it can be further refined to a totally ranked hierarchy. However, Boersma (ibid) shows that under the assumption of pooling ties the non-existence of at least one total ranking inconsistent with the language data is not guaranteed. Suppose that the initial state grammar of the hypothetical learner is as shown in Tableau (2-11) below, where o_1 is the intended winner.

(2-11) *The learner's optimal candidate in the initial state: EDCD with pooling ties* (from Boersma 2008:4)

i_1	C_1	C_2	C_3
✓☞ o_1			*
o_2	*	*	

According to EDCD, the learner has to compute a loser, which is the optimal form according to her current grammar. It just so happens that in this case the loser equals the winner. Since no demotions are necessary to account for the target language data, the convergence criterion is met. EDCD stops, having produced the following stratified grammar:

(2-12) *Final stratified hierarchy obtained by EDCD* (from Boersma 2008:5)

$\{C_1, C_2, C_3\}$

Such a stratified hierarchy is a correct grammar of a target language in a sense that it is consistent with all data. However, it needs to be further refined to a totally ranked hierarchy in accordance with the *strict domination* principle (see Prince & Smolensky 1993). According to Tesar and Smolensky (1993:11), the stratified hierarchy that is the output of RCD algorithm "represents a class of all totally-ranked constraint hierarchies which give rise to the target language *L*", and therefore "the same optimal outputs arise regardless of the ranking of the [...] constraints". However, as pointed out by Boersma (ibid), this claim does not hold true of the stratified hierarchy above, as the permutations of the given three constraints, *inter alia*, give rise to the grammars $C_3 >> C_1 >> C_2$ and $C_3 >> C_2 >> C_1$, which incorrectly predict the sub-optimal candidate o_2 to win. On the basis of this illustration, Boersma (2008:5) concludes that EDCD with pooling ties is not guaranteed to converge on a correct totally ranked hierarchy. Boersma (2004:7) further notes that the problem results from the inaccessibility of the informative loser o_2, which can never be computed as an optimal output of the learner's grammar given the violation profile as in Tableau (2-11) and the assumption of pooling ties.

In order to solve this problem, Boersma (2008:8) proposes to discard the pooling ties assumption, and adopt so called *permuting ties* instead (with a reference to Antilla 1997). According to Boersma (ibid), in order to assess a relative harmony of competing candidates under the assumption of permuting ties, the learner randomly chooses a total ranking consistent with the current stratified hierarchy, rather than collapsing all unranked constraints within a stratum together. Boersma (ibid) calls this approach "Variationist EDCD".

(2-13) *The learner's optimal candidate in the initial state: Variationist EDCD* (from Boersma 2008:8)

i_1	C_1	C_2	C_3
✓☞ o_1			*
☞o_2	*	*	

Consider Tableau (2-13), for example. The stratified hierarchy illustrated by the Tableau is consistent with six total rankings. Every time the learning algorithm receives a learning datum (an adult output), the learner will randomly choose one of such rankings and compute the optimal output of his grammar under such ranking. If the optimal output of the learner's grammar equals adult output, no learning takes place. However, when the learner encounters the same datum again, she chooses another random total ranking, let it be $C_3 >> C_1 >> C_2$. Under such ranking the output of the learner's grammar is a sub-optimal candidate o_2. This is illustrated in the Tableaux below.

(2-14) *Error-driven learning with permuting ties* (from Boersma 2008:8)

i_1	C_1	C_3	C_2
✓☞ o_1		*	
o_2	*!		*

i_1	C_3	C_1	C_2
✓ o_1	*!		
☞o_2		*	*

When the error has been detected, the learner gets evidence to demote the offending constraint to the lower stratum. The full learning procedure for Variationist EDCD is given below.

(2-15) *Learning procedure for Variationist EDCD* (from Boersma 2008:14)

1. The learner receives an input-output pair (*i*, *o*).
2. The learner determines her own optimal output, given the input *i*:

   2a. The learner randomly chooses a total constraint ranking consistent with her current stratified ranking under the assumption of permuting ties.
   2b. The learner determines the outputs that are optimal under this total ranking (there may be multiple optimal outputs, if two candidates have identical violation patterns).
   2c. The learner randomly chooses her output from the set of optimal outputs determined in 2b.

3. If the learner's own output for *i* is different from *o*:

   3a. Determine the stratum *s* that contains the highest-ranked constraint that prefers *o* over the learner's own output.
   3b. All constraints that prefer the learner's own output over *o* and that are not already ranked in a lower stratum than *s*, are demoted into the stratum just below *s*.

According to Boersma (2008), Variationist EDCD is guaranteed to converge on a totally refinable stratified hierarchy. However, it has been recognized (see Tessier 2006, Stemberger & Bernhardt 2001) that, although efficient and restrictive, error-driven constraint demotion algorithms are not suitable to model a life-like learning situation. The problem is that according to error-driven algorithms (Original EDCD, Variationist EDCD, BCD), the learner is able to fix her production error as soon it has been detected, by demoting the responsible constraint by one fell swoop. However, early production data (see Tessier 2006, Gnanadesikan 1995, Smith 1973) show that on their way to the correct adult pronunciation children often go through one or several intermediate stages, where their faulty patterns are but partially fixed. For example, Tessier (2006:409) (with reference to Rose 2000), provides some production data from a longitudinal study of Québécois French learners trying to acquire complex onsets. According to the data, the learning process could be subdivided into three stages: the initial stage of acquisition was characterized by cluster reduction, during the intermediate stage the learner retained complex onsets of stressed syllables only, and in the final stage the child learned to produce all complex onsets faithfully. As noted by Tessier 2006, error-driven algorithms do not predict the existence of such intermediate stages. Both Tessier 2006 and Stemberger & Bernhardt 2001 recognize that error-driven algorithms should be modified in order to account for the 'gradualness' of acquisition. According to Stemberger & Bernhardt 2001, "it is necessary to adopt a variant [of EDCD] in which all changes are small".

Actually, such a variant is readily available. It is called Minimal Gradual Learning Algorithm, or Minimal GLA (see Boersma 1998, Boersma 2008). The only thing that distinguishes Minimal GLA from Variationist EDCD is that Minimal GLA can demote only one constraint at a time, and by only one stratum at a time. This predicts that the learning is gradual, i.e. the learner is no longer able to fix her production errors in one fell swoop.

Summarizing our discussion so far, we can list the crucial requirements any good learning algorithm should meet.

(2-16) *Requirements to learning algorithms*

The learning algorithm should:

a. converge upon a totally refinable stratified ranking;
b. derive a maximally restrictive ranking compatible with the learning data;
c. account for empirical acquisition data (e.g. early production/perception data)

To the best of our judgement, Minimal Gradual Learning Algorithm (as formulated in Boersma 2008), satisfies all the abovementioned requirements. Therefore, our further discussion of acquisition will be based upon this algorithm, and compatibility to this algorithm will be one of the criteria against which we will evaluate the proposed modifications of Optimality Theory.

2.3 Learning Underlying Representations

In the previous section we have considered different learning algorithms that enable the learner to acquire the constraint ranking that defines the set of permissible outputs of the target language. But, as noted by Kager (1999a:222), "much more is at stake in learning a grammar than constraint ranking".

So far we have based our discussion on the simplifying assumption that both surface forms and correct corresponding underlying forms are readily available for the language learner. In real life, however, the learner faces a complex task of inferring both the underlying representations and the constraint ranking solely on the basis of positive evidence in the form of adult outputs. In this section we will consider different mechanisms that have been proposed to account for the acquisition of lexical representations.

According to Tesar (2007:572), at least two different strategies compatible with the above-discussed learning algorithms have been proposed to deal with this problem. The solution proposed by Tesar & Prince 2003 involves testing different hypothesized underlying forms against a phonotactic ranking constructed by BCD during the initial stage of acquisition assuming the identity map. The alternative solution suggested by Tesar et al. 2003 uses the inconsistency detection mechanism to choose the correct underlying form, thus enabling the learner to acquire correct underlying forms simultaneously with the constraint ranking.

2.3.1 Using phonotactics to learn phonological alternations

It has been assumed (see, *inter alia*, Prince & Tesar 2004, Tesar & Prince 2003) that at the initial stage of language acquisition the learners are incapable of morphological analysis and therefore treat each word separately, as if it were morphologically unrelated to other words. It is also generally believed (Smolensky 1996a, Kager 1999a, Prince & Tesar 2004, Tesar & Prince 2003) that at the early stage of acquisition the learners take the lexical representation of every word to be identical to its surface representation as produced by adults. The proof that such a view is in fact warranted comes from several empirical studies of early child production (see Smith 1973, Gnanadesikan 1995).

Given these assumptions about the learner's underlying forms and her ignorance of morphology, at the initial state of learning the crucial task of algorithms like BCD is to establish a ranking that would map each 'legal' input form to itself, while at the same time mapping 'illegal' inputs to 'legal' outputs. As we have seen in the previous section, there are at least two algorithms that can handle the task.

Later in acquisition, however, learners become aware of morphology and, consequently, of alternations. They realize that the same morpheme has different surface forms in different contexts. To illustrate how the learner's knowledge about the target language expands over time, consider the tables below (from Tesar & Prince 2003).

(2-17) *Morphologically opaque forms of language D* (adapted from Tesar&Prince 2003:11)

Solid Lexical Unit
tat
dat
tate
tade
date
dade

(2-18) *The fully segmented forms of language D* (from Tesar&Prince 2003:13).

Bare Root	Root + Suffix
tat_1	tad_1-e_5
tat_2	tat_2-e_5
dat_3	dad_3-e_5
dat_4	dat_4-e_5

The observation that the same morpheme may have different surface forms depending on the context warrants the departure from the identity map, and at this point assigning correct underlying representations to alternating items becomes crucial.

According to Tesar & Prince (2003:13), though the learner realizes that assigning a single underlying form that would be identical to both surface instances of an alternating morpheme is impossible, she still adheres to the identity map as closely as possible. In other words, the learner assumes that correct underlying representation of non-alternating morphemes equals their surface representation. Then, the learner determines the invariant features of alternating morphemes (here: everything except for the voicing feature of the final obstruent) and fixes them in the underlying form. Having done that, the learner creates several possible underlying representations for each alternating morpheme, differing only in the value of the alternating feature. Hypothesized lexical representations for the surface forms from Table (2-18) are given below.

(2-19) *Underlying form hypotheses for the morphemes of the paradigm* (from Tesar&Prince 2003:14).

Morpheme	**UF Hypotheses**
#1	/tat/$_1$, /tad/$_1$
#2	/tat/$_2$
#3	/dat/$_3$, /dad/$_3$
#4	/dat/$_4$
#5	/-e/$_5$

These hypothesized underlying forms are now to be tested against the constraint ranking constructed by the learner during the phonotactic learning stage. Suppose that the ranking our learner arrived at is as in (2-21), with the constraints defined in (2-20).

(2-20) *Definitions of constraints active in language D* (from Tesar&Prince 2003:3)

NoVoi	no voiced obstruents
NoSFV	no syllable-final voiced obstruents
IVV	no inter-vocalic voiceless obstruents
IDVoi	surface voicing must match underlying voicing

(2-21) *Grammar of language D established by phonotactic learning* (from Tesar&Prince 2003:10)

NoSFV >> IDVoi >> {NoVoi, IVV}

Now the learner has to construct inputs for each of the words containing the alternating morpheme. In case of morpheme #1, two underlying form hypotheses give rise to two sets of inputs: /tat/ for [tat] and /tat+e/ for [tade], and /tad/ for [tat] and /tad+e/ for [tade]. Note also that choosing /tat/ as an underlying form means treating tat～tade alternation as intervocalic voicing, while choosing /tad/ as an underlying form means treating the same alternation as syllable-final obstruent devoicing.

According to Tesar & Prince (2003:15), at this point the learner does not attempt to construct any winner-loser pairs. She just checks if her grammar produces correct results when fed newly-constructed inputs. Only if the phonotactic ranking is insufficient will the learner be motivated to construct winner-loser pairs. Tableau (2-22) below shows how output candidates corresponding to different inputs fare with respect to the learner's current grammar.

(2-22) *Testing hypothesized underlying representations against phonotactic ranking*

	NoSFV	**IDVoi**	**NoVoi**	**IVV**
/tat/				
☞tat				
tad	*!	*	*	
/tat+e/				
💣tate				*
☹tade		*!	*	
/tad/				
☞tat		*		
tad	*!		*	
/tad+e/				
tate		*!		*
☞tade			*	

The fact that given the underlying form /tad/ for the alternating morpheme [tat] the grammar correctly maps underlying forms to surface forms in both environments, prompts the learner to choose /tad/ as a correct underlying representation of the alternating morpheme [tat].

In the simplified case considered above, the only possible strategy to fix the marked input form was to change the voicing specification of the final obstruent. Tesar & Prince (2003:17) note that in cases where several repair strategies are available, pure phonotactics are not enough to determine the constraint ranking of the language and correctly assign underlying forms. To illustrate their point, they adopt segment deletion as an alternative way of avoiding marked voiced final obstruents. The ranking is now as follows (from Tesar & Prince 2003:18):

(2-23)

NoSFV >> {IDVoi, Max} >> {NoVoi, IVV}

Under such a ranking, the output candidates corresponding to the underlying form /tat/ will fare as in Tableau (2-22) above, with [tate] incorrectly predicted to be optimal. For the underlying /tad/ the situation is somewhat more complex, and this is illustrated in Tableau (2-24) below.

(2-24)

	NoSFV	**IdVoi**	**Max**	**NoVoi**	**IVV**
/tad/					
tad	*!			*	
☞tat		*			
💣ta			*		
/tad+e/					
☞tade				*	
tate		*!			*
tae			*!		

As shown in Tableau (2-24), outputs [tat] and [ta] tie. This means that the learner cannot map an underlying form to a single output form. The learner attributes the error to the deficiency of her current constraint ranking. The learner then constructs new winner-loser pairs (in addition to phonotactic winner-loser pairs), and applies BCD in the usual way to derive a new ranking. The set of winner-loser pairs with violation pattern for the input /tat/$_1$ is given in (2-25) below.

(2-25) *Inconsistent winner-loser pairs for /tat/$_1$* (adapted from Tesar&Prince 2003:20)

	Lexicon	**Winner ~ Loser**	**NoVoi**	**NoSFV**	**IVV**	**IDVoi**	**Max**
(a)	/tate/	tate ~ tae			L		W
(b)	/tate/	tate ~ tade	W		L	W	
(c)	/tat/$_1$, /-e/$_5$	tad-e ~ tat-e	L		W	L	

In accordance with BCD, the learner will rank NoSFV in the top-stratum, since it is unviolated by either winners or losers. Then the learner will rank Max, since it only prefers winners. The remaining three constraints cannot be ranked because each of them prefers one loser. The learner concludes that the set of winner-loser pairs is inconsistent. According to Tesar & Prince (2003:20), the learner assumes that the problem lies in the underlying form used. The learner constructs winner-loser pairs corresponding to the alternative underlying form /tad/$_1$.

(2-26) *Consistent winner-loser pairs for /tad/$_1$* (adapted from Tesar&Prince 2003:20)

	Lexicon	**Winner ~ Loser**	**NoVoi**	**NoSFV**	**IVV**	**IDVoi**	**Max**
(a)	/tate/	tate ~ tae			L		W
(b)	/tate/	tate ~ tade	W		L	W	
(c)	/tad/$_1$	tat ~ ta				L	W

This set of winner-loser pairs is consistent and allows the learner to establish the correct ranking of constraints whereby Max dominates IDVoi.

Thus by effectively using the ranking established during pure phonotactic learning, the learner managed to assign the correct underlying representation to the alternating morpheme and to establish the constraint ranking inducing the language-specific repair strategy for marked inputs.

2.3.2 Surgery in Language Learning

The approach to learning underlying forms adopted by Tesar, Alderete, Horwood, Merchant, Nishitani & Prince 2003 is somewhat different. Tesar *et al.* 2003 proceed from the assumption that at the initial state the learner has already mastered the morphology of the target language. The input into the learner's grammar, therefore, consists of fully morphologically segmented surface forms. Then the learner assigns a hypothesized underlying form to each input. This is done in the following fashion: the learner assumes that for non-alternating morphemes the underlying form equals surface representation, while alternating forms are assigned a 'default' feature value. When the underlying representations have been assigned, the learner constructs winner-loser pairs for each input. BCD is applied to the set of winner-loser pairs. If the BCD detects inconsistency, i.e. if there is no such ranking that would render the correct outputs, the learner realizes that her hypothesized lexical representations have to be changed. The learner changes a previously hypothesized default feature value in the underlying representation of alternating morphemes, one morpheme at a time, until the inconsistency is resolved. If changing the underlying representation of some morpheme did not resolve the inconsistency, such change is revoked. The change resulting in resolution of inconsistency is retained.

A change in underlying representation has important consequences for the set of winner-loser pairs previously constructed and stored by the learner. Tesar et al. propose that instead of discarding ‘obsolete’ winner-loser pairs altogether, the learner makes sure that the relevant-winner loser pairs get adjusted to the new underlying representation (the process of adjustment here is called *surgery*). When no more adjustments are needed to extend to the new data, the learner has converged on the most restrictive ranking capturing the data of the target language, at the same time having correctly assigned underlying representations to alternating and non-alternating morphemes.

2.3.3 Residual Issues: the Free Ride

For the purposes of both algorithms considered above it was assumed that learners adhere to identity map in case of non-alternating morphemes, i.e. if the morpheme has the same surface form in all contexts, this surface form is taken to be identical to its underlying form. However, according to McCarthy 2004, in certain cases the departure from the identity map is warranted for non-alternating items as well. To support his proposal, McCarthy (2004:3) provides data from Sanskrit (with reference to de Haas 1988, Gnanadesikan 1997, Schane 1987, Whitney 1989), where surface long mid vowels [e:] and [o:] are derived by coalescence from /ai/ and /au/. There are, however, instances of surface [e:] in non-alternating morphemes.

(2-27) *Sanskrit coalescence* (from McCarthy 2004:3)

/tava indra/	tave:ndra	‘for you, Indra (voc.)’
/hita upadaiʃah/	hito:pade:ʃah	‘friendly advice’

According to McCarthy 2004, the evidence from alternations will make the learner conclude that some instances of [e:] are derived from underlying /ai/. While the adherence to the identity map will force them to posit underlying /e:/ for non-alternating morphemes. In terms of constraint ranking, it will force the learner to rank the faithfulness constraint demanding identity to underlying vowel height above the markedness constraint prohibiting mid vowels. Such a grammar, however, is not the most restrictive grammar of Sanskrit, since it also permits surface [e], which does not occur in the language. According to McCarthy 2004, in order to acquire the most restrictive grammar of Sanskrit, the learner has to extend her observation that surface [e:] derives from underlying /ai/ to non-alternating forms as well. Stated in general terms, McCarthy's 2004 proposal is the following:

(2-28) *The Free Ride* (from McCarthy 2004:11)

> [L]earners, whenever alternations lead them to discover a new unfaithful map, always attempt to generalize that map [...] across the entire language.

In other words, if at some point in morphophonemic learning the learner detects that some instances of surface [B] are derived from underlying /A/, the learner assumes that all instances of surface [B] are derived from underlying /A/. If such a hypothesis leads to inconsistencies in ranking (i.e. if there is no ranking that can capture the data), the hypothesis is discarded.

2.4 Summary

Above we have seen that the algorithms proposed by Tesar & Prince 2003 and Tesar *et al.* 2003 both successfully solve the task of learning underlying representations. While we realize that further empirical testing is necessary to judge which of the proposed algorithms reflects the real-life learning situation more appropriately, the idea that at the early stages of acquisition the learners are oblivious to morphology seems plausible to us. Therefore, for the purposes of our further discussion we will use the algorithm proposed by Tesar & Prince 2003, whereby the learners go through the stage of pure phonotactic learning before they attempt morphological segmentation of output forms.

Chapter 3: Opacity in Optimality Theory

3.1 Introduction

The term 'opacity' is used to refer to the phenomenon whereby a surface form of some language L has exceptionally undergone or failed to undergo a certain phonological process active in language L. According to McCarthy 1999, there can be distinguished two basic types of opacity:

(3-1) *Types of opacity* (from McCarthy 1999:2)

a. Linguistically significant generalizations are often not surface-true. That is, some generalization G appears to play an active role in some language L, but there are surface forms of L (apart from lexical exceptions) that violate G. Serialism explains this by saying that G is in force at only one stage of the derivation. Later derivational stages hide the effect of G, and may even contradict it completely.

b. Linguistically significant generalizations are often not surface-apparent. That is, some generalization G shapes the surface form F, but the conditions that make G applicable are not visible in F. Serialism explains this by saying that the conditions on G are relevant only at the stage of the derivation when G is in force. Later stages may obliterate the conditions that made G applicable (e.g., by destroying the triggering environment for a rule).

In case of 3-1(a), the process fails to apply in the expected context, while in 3-1(b) the process applies outside the expected context. Therefore, to refer to these two types of opacity McCarthy (1999) uses the terms *underapplication* and *overapplication* respectively. Schematically, the above-given definition can be represented as follows (from Baković 2007:2; Kiparsky 1971):

(3-2)

A process **P** of the form *A*→ *B* / *C_D* is *opaque*
to the extent that there are surface representations of the form:

a. A in the environment C_D, or *[=non-surface-true/underapplication opacity]*

b. B derived by P in the environment other than C_D *[=non-surface-apparent/overapplication opacity]*

Serialism views the grammar as consisting of ordered rules, where the output of one rule is the input of the following rule. According to Baković (2007:2), in serialism opacity is said to arise from some rule **Q** ordered after the rule corresponding to the process **P**. Overapplication corresponds to the situation where the rule **Q** destroys the context for the process **P** to apply (i.e. **Q** *bleeds* **P**), but because **Q** is ordered after **P** (i.e. the rule order is *counterbleeding*), the generalization expressed by **P** is surface-true, but the reasons for **P** to apply are not apparent in the surface form. Underapplication, in its turn, corresponds to the situation where the rule **Q** creates the context for **P** to apply (i.e. **Q** *feeds* **P**), but because **P** is ordered before **Q** (i.e. the rule order is *counterfeeding*) the generalization expressed by **P** is not surface-true.

Thus, the serialist account crucially relies on the existence of some intermediate form that deviates from the input and from the actual output. According to Prince & Smolensky (1993:6), in 'classic' Optimality Theory the "Input → Output map has no internal structure: all possible variants are produced by Gen in one step and evaluated in parallel". Precisely this property makes opacity a challenging issue for OT.

It should also be noted that opaque generalizations are by no means limited to fully-developed grammars. There is an extensive empirical base (see, *inter alia*, Smith 1973, Dinnsen *et al*. 2000, Dinnsen 2008) that demonstrates that both counterbleeding and counterfeeding are also common for developing child grammars. To distinguish between these two types of opacity, we will use the terms 'emergent', 'spontaneous', 'developmental' or 'non-target-like' to refer to the opacity effects in child grammars and the term 'target-like' to refer to the opacity in fully-developed grammars (the terms are due to Wolf 2008, Dinnsen *et al*. 2000, Jesney 2005 respectively).

This chapter is organized as follows: in Section 3.2 we will look at different types of opacity and illustrate why they posit a challenge for classic OT. In Section 3.3, we will look at different OT-based approaches aimed at facilitating the analysis of opaque generalizations. In Subsection 3.3.1 we will deal with the approaches that do so by enriching representational complexity and the constraint component Con, in Subsection 3.3.2 we will consider approaches that posit intermediate derivational stages, and in Subsection 3.3.3 we will discuss approaches based on the Horizontal Correspondence principle. In Section 3.4 we will briefly summarize the results of our discussion.

3.2 Opacity in Optimality Theory

According to Bermúdez-Otero (forthcoming), the failure of classic OT to account for opaque generalizations results from OT being *output-oriented*.

(3-3) *Output Orientation* (from Bermúdez-Otero forthcoming:5)

In any grammatical mapping,

a. constraints evaluate either the structure of output candidates (markedness) or the relationship between output candidates and other grammatical representations (correspondence);
b. all output candidates are evaluated in parallel.

Below we will consider real-life situations where output orientation hinders classic OT from yielding satisfactory analysis of opaque generalizations. First we will look at the case of counterbleeding opacity in Yokuts, and then we will consider two types of counterfeeding opacity in Bedouin Arabic.

3.2.1 Counterbleeding

As we already know, the term 'counterbleeding opacity' refers to the cases when the reason for the application of some phonological process P is not obvious on the surface form. In serialist terms, it is said that in such a case the process P is rendered opaque by the later application of some process Q. According to McCarthy (1999:25), counterbleeding opacity can be schematically represented as follows:

(3-4) *Non-Surface-Apparent or Counter-Bleeding Opacity* (from McCarthy 1999:25)

UR	ABC#
B→D/_C	ADC#
C→E/_#	ADE#
SR	ADE#

On the scheme above, the context of the first process was destroyed by the application of the second process. Therefore, if we disregard intermediate stages of derivation, the first process will appear to have applied out of context, i.e. /B/ appears to turn into [D] before [E].

Let's now consider a real-life example of counterbleeding opacity. McCarthy (1999:22) presents a case of Yokuts language, where the interaction of long vowel lowering and closed syllable shortening gives rise to non-surface-apparent generalization.

(3-5) *Yokuts Vowel Alternations* (from McCarthy 1999:22)

a. Vowels are shortened in closed syllables:

/pana:/	panal	cf. pana:hin	'might arrive/arrives'
/hoyo:/	hoyol	cf. hoyo:hin	'might name/names'

b. Long high vowels are lowered:

/ʔili:/	ʔile:hin		'fans'
/c'uyu:/	c'uyo:hun		'urinates'

c. Vowels shortened in accordance with (a) are still lowered:

/ʔili:/	ʔilel		'might fan'
/c'uyu:/	c'uyol		'might urinate'

In a serialist model, the Yokuts data can be captured by counterbleeding order of shortening and lowering rules. Consider the following:

(3-6) *Yokuts Serial Derivation* (from McCarthy 1999:22)

UR	/ʔili:-l/
Lowering	ʔile:l
Shortening	ʔilel

An optimality-theoretic analysis of Yokuts counterbleeding is far from being straightforward. According to Baković 2007 (also see McCarthy 1999), the OT analysis of each individual process is unproblematic. Thus, according to Baković, long vowel lowering is due to a high-ranking markedness constraint against long high vowels, No-Long-High, ranked above the faithfulness constraint Ident(high). To rule out a shortening scenario, a faithfulness constraint Max-μ should also dominate Ident(high). This is illustrated in the tableau below.

(3-7) *Long vowel lowering* (from Baković 2007:7)

/ʔili:+hin/	No-Long-High	Max-μ	Ident(high)
a. [ʔili:hin]	*!		
☞b. [ʔile:hin]			*
c. [ʔilihin]		*!	

According to Baković (2007:7), the process whereby long vowels are shortened in closed syllables is due to the markedness constraint No-Long-Closed dominating Max-μ. This is illustrated in the tableau below.

(3-8) *Closed syllable shortening* (from Baković 2007:7)

/pana:+l/	No-Long-Closed	Max-μ
a. [pana:l]	*!	
☞b. [panal]		*

The interaction of the two abovementioned processes is a principled problem for OT. Given the constraint ranking established so far, the input containing a long high vowel in a closed syllable can be ‘repaired’ by simply shortening the vowel. The output candidate with a shortened high vowel in a closed syllable satisfies both high-ranking markedness constraints while violating Max-μ. The intended winner, however, contains a shortened *and* lowered vowel, thus incurring an unmotivated violation of Ident(high). This is illustrated in Tableau (3-9) below.

(3-9) *Failure of counterbleeding in OT* (from Baković 2007:7)

/ʔili:+l/	No-Long-High	No-Long-Closed	Max- μ	Ident(high)
a. [ʔili:l]	*!	*!		
b. [ʔile:l]		*!		*
💣c. [ʔilil]			*	
☹d. [ʔilel]			*	*!

In Tableau (3-9), the intended winner in (d) is harmonically bounded by candidate (c). That is, there is no such constraint unviolated by the intended winner that would be violated by candidate (c). Therefore, there is no such ranking that would correctly predict candidate (d) to win. In subsequent sections we will consider some possible solutions for the problem outlined here. But before, let us consider another type of opaque generalizations, namely, counterfeeding opacity.

3.2.2 Counterfeeding

Recall from Section 3.1 that the term ‘counterfeeding opacity’ refers to the cases when the reasons for non-application of some process P are not obvious from the surface form. According to McCarthy (1999:31), counterfeeding opacity can be of two types, namely, *counterfeeding on environment* and *counterfeeding on focus*. In serialist terms, in the first case the generalization is not-surface true because the process’s environment was met too late in the derivation; in the second case, the generalization is not surface-true because the segment to be affected by the process was introduced too late in the derivation. Schematically, it can be illustrated as follows:

(3-10) *Non-Surface-True or Counter-Feeding Opacity* (from McCarthy 1999:31)

a. Counter-Feeding on Environment

UR	ABC
B→D/_E	does not apply
C→E/_#	ABE

b. Counter-Feeding on Focus

UR	ABC
D→E/A_	does not apply
B→D/_C	ADC

Let's now consider the real-life examples of counterfeeding opacity. In Bedouin Arabic, the interaction of phonological processes gives rise to both counterfeeding on environment and counterfeeding on focus.

(3-11) *Phonological alternations in Bedouin Arabic* (adapted from McCarthy 2007)

a. Short high vowels are deleted from non-final open syllables

/kit**i**b-at/	[ˈkitbat]	'it (m.) was written'
/ʃar**i**b-at/	[ˈʃarbat]	'she drank'

b. Short low vowels become high in non-final open syllables

/katab/[1]	[kitab]	'he wrote'
/kabak/	[kibak]	'cufflink'

c. Epenthesis applies to break up final consonantal clusters

/gabr/	[gabur]	'grave'

In example 3-11 (a) we can see that in Bedouin Arabic short-high vowels are normally deleted from non-final open syllables. However, as shown in 3-11(b), if a high vowel is the result of low-vowel raising, no syncope takes places. Raising, in its turn, applies to short low vowels in open syllables (as in 3-11(b)), but fails to apply if an open syllable is created by epenthesis (as in 3-11(c)).

---

[1] See McCarthy (2007:189-191) for the discussion about the underlying representation of [kitab]

In rule-based phonology, opaque forms of Bedouin Arabic are the result of the counterfeeding order of the rules that express generalizations about the phonological processes active in the language. Consider the following:

(3-12) *Two types of underapplication opacity in Bedouin Arabic* (adapted from Baković 2007:5)

a. counterfeeding-on-focus

UR	/katab/	
i→Ø/_σ	n/a	
a→i/_σ	/kitab/	
SR	[kitab]	‘he wrote’

b. counterfeeding-on-environment

UR	/gabr/	
a→i/_σ	n/a	
Epenthesis	/gabur/	
SR	[gabur]	‘grave’

Below we will illustrate that the abovementioned cases of counterfeeding opacity are problematic for OT in its classical formulation[2]. Following McCarthy (2007:103), we assume that the i→Ø alternation is due to a markedness constraint against high vowels in open syllables *iCV dominating Max.

(3-13) *Syncope in Bedouin Arabic*

/ʃa**ri**b-at/	*iCV	Max
a. [ʃa**ri**bat]	*!	
☞b. [ʃarbat]		*

Following McCarthy (1999:7), we also assume that the a-raising process is due to the markedness constraint *aCV ranked above Id(low).

(3-14) *Raising in Bedouin Arabic*

/kabak/	*aCV	Id(low)
a. kabak	*!	
☞b. kibak		*

---

[2] See Baković 2007, McCarthy 2007 and references therein for the analyses based on extending the inventory of faithfulness constraints. Also see Chapter 3 for the discussion of chain-shifts in OT-CC.

In Bedouin Arabic, syncope only affects underlying high vowels. High vowels that are the product of low-vowel raising always surface faithfully. Serialist models can capture this generalization by ordering syncope before raising, as illustrated above in 3-12 (a). In 3-12 (a), the output of the syncope rule is the input to the raising rule. In OT, where ordered rules are replaced by ordered constraints, all output candidates are evaluated in parallel. Transparent application of syncope requires that the *iCV constraint dominates Max. This ranking, however, makes incorrect predictions in case of counterfeeding on focus.

(3-15) *Counterfeeding-on-focus in Bedouin Arabic*

/katab/	*iCV	*aCV	Max	Id(low)
a. katab		*!		
☹b. kitab	*!			
💣c. ktab			*	

In Tableau (3-15) above, the faithful candidate (a) is eliminated by the high-ranked markedness constraint that bans low vowels in open syllables. The opaque form in (b) incurs a violation of high-ranked *iCV. Consequently, the transparent form *[ktab] is incorrectly selected as optimal, since it only violates the low-ranking Max constraint. This situation cannot be fixed simply by ranking *iCV below Max, because such a ranking would prevent the normal application of syncope in mappings like /ʃa**ri**bat/ → [ʃarbat].

Counterfeeding-on-environment is equally problematic. In Bedouin Arabic, a-raising process does not apply if its context is created by epenthesis. In a serialist model this can be achieved by ordering epenthesis after a-raising. Classic OT machinery does not allow for this possibility. Following McCarthy 2007, we assume that vowel epenthesis is due to the markedness constraint against final consonant clusters *Comp-Coda being ranked above Dep. Normal application of a-raising requires that *aCV dominates Id(low). Again, OT ranking fails to capture the opaque generalization.

(3-16) *Counterfeeding-on-environment in Bedouin Arabic*

/gabr/	*Comp-Coda	*aCV	Dep	Id(low)
☹a. gabur		*!	*	
💣b. gibur			*	*
c. gabr	*!			

In Tableau (3-16) above, faithful candidate (c) is ruled out by the high-ranked well-formedness constraint *Comp-Coda that expresses the prohibition against consonantal clusters in coda position. The intended winner, opaque form [gabur] crucially violates high-ranked *aCV. Therefore, the transparent form *[gibur] is incorrectly chosen as optimal. We cannot solve this problem by ranking *aCV below Id(low), because this ranking would preclude the normal application of a-raising.

As we have seen in this section, classic OT fails to capture opaque generalization of both the counterbleeding and counterfeeding type. As we have noted in the beginning of the section, this failure is probably due to the output orientation of classic OT. There have been several attempts to incorporate opacity in OT. Some approaches do so by eliminating the output orientation condition, some others by enriching underlying representations or the constraint component Con. In the next section we will consider different approaches to the problem of opacity in OT and briefly discuss their relative drawbacks and merits.

3.3 Approaches to Opacity in Optimality Theory

According to McCarthy (2007), attempts to incorporate opacity in classic OT fall into four broad categories. Our discussion in this section will be loosely based on the classification identified by McCarthy (2007).

(3-17) *Approaches to opacity in OT* (from McCarthy 2007:27)

a. Changes in substantive properties of phonological representation or the constraint component Con. The goal is to analyze some or all cases of opacity by enriching representations or creating new constraints.

b. Introduction of intermediate derivational stages and something like rule ordering to OT.

c. Introduction of an equivalent of intermediate derivational stages, but without any direct counterpart to rule ordering.

d. Reinterpretation of opacity as a mechanism for preserving underlying contrasts.

The approaches to opacity to be considered in this section can be roughly subdivided into two large (and partly overlapping) groups: those that rely on expanding some basic assumptions about the nature of underlying representations or the constraint component Con and those that make reference to forms distinct from the input and the actual output. We will see that the latter approaches prove to be more successful at least from the typological perspective. However, as it will shortly become obvious, even those approaches that make correct predictions about the typology of opaque generalizations face a lot of difficulties with respect to the acquisition challenge.

3.3.1 Enriching Representations and Creating New Constraints

Approaches to be considered in this section remain faithful to the output orientation principle inherent in classic OT. The path that they follow is the one of expanding the notion of underlying representations and the constraint set.

3.3.1.1 Containment and Turbidity

Prince & Smolensky (1993) analyze two cases of opacity, counterfeeding in Lardil and counterbleeding in Fula (also see a detailed discussion in McCarthy 2007). The analysis offered by Prince & Smolensky (1993) crucially depends on their interpretation of the segmental deletion process. In accordance with the principle of Containment (term due to McCarthy & Prince 1993:21), input segments are never completely deleted from the representation, while deletion effects result from three additional assumptions:

(3-18) *Containment* (from McCarthy 2007:28)

(i) Underlying representations lack prosodic structure, particularly syllabification.
(ii) Phonological material may remain unincorporated into prosodic structure.
(iii) Unincorporated phonological material receives no phonetic interpretation.

According to Prince & Smolensky (1993:109), in Lardil only a limited number of consonants can appear in coda position. Other consonants are deleted from coda position. There is also another process (Prince & Smolensky 1993:108), whereby word-final vowels are deleted in the nominative case. These two processes interact, i.e. whenever a deletion of a word-final vowel results in a disallowed coda consonant, such a consonant is deleted as well. Crucially, coda deletion does not feed final vowel deletion, i.e. if coda deletion results in a word-final vowel in the nominative, such vowel is allowed to surface. Consider the following:

(3-19) *Interaction of Coda Condition and vowel deletion in Lardil* (adapted from Prince& Smolensky 1993:110)

UR	/ŋawuŋawu/		/muřkunima/	
Vowel Deletion	/ŋawuŋaw/		/muřkunim/	
Coda Condition	/ŋawuŋa/		/muřkuni/	
SR	[ŋawuŋa]	‘termite’	[muřkuni]	‘nullah’

The process of vowel deletion is non-surface-true: it seems to have underapplied in words like [ŋawuŋa] and [muřkuni]. However, in accordance with the principle of Containment, the representation of words *ŋawuŋa* and *muřkuni* looks like [ŋawuŋa<w><u>] and [muřkuni<m><a>] respectively. Vowel deletion does not apply because its structural description is not met: technically, the vowels [a] and [i] in these words are not final.

In Fula (Prince & Smolensky 1993:235), the prohibition on continuant geminates triggers ‘hardening’ (the term is due to McCarthy 2007:29) of [+continuant] segments. In addition, the language also prohibits sequences of long vowels and geminates, which results in degemination. The two processes interact in counterbleeding order, which gives rise to surface forms where hardening seems to have overapplied. Consider the following:

(3-20) *Interaction of hardening and degemination in Fula* (adapted from Prince& Smolensky 1993:235)

UR	/la:w:i/
Hardening	/la:b:i/
Degemination	/la:bi/
SR	[la:bi]

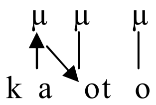
The solution offered by Prince & Smolensky (1993:ibid) is very similar to their treatment of Lardil counterfeeding, and is based on the assumption that deletion means ‘underparsing’. A markedness constraint against continuant geminates is violated regardless of whether such geminates are parsed or not. The constraint against sequences of long vowels and geminates, in its turn, is only violated if geminates are syllabified. Thus, the ‘transparent’ candidate [la:w<w>i] violates the constraint against continuant geminates, while the ‘opaque’ candidate [la:b<b>i] satisfies both markedness constraints.

According to McCarthy (2007:29-30), certain cases of opacity don't lend themselves to Containment-based analysis. Such is the case of Bedouin Arabic, where epenthesis counterfeeds low vowel raising in open syllables, resulting in /gabr/ → [gabur] mapping. Since underlying representations have no prosodic structure, there is no such stage of syllabification where /a/ is in a closed syllable and therefore protected from raising. On the other hand, as McCarthy (ibid) notes, certain transparent interactions become problematic if analyzed from the point of view of Containment. Thus, in Maltese syncope feeds voicing assimilation, producing mappings like /ni-kti-bu/ → [‘nigdbu]. Under containment, the underlying representation of [‘nigdbu] is [‘nigd<i>bu], which is implausible, since voicing assimilation has never been observed to apply across a pronounced vowel. An assumption that the unparsed vowel has no effect on determining segment adjacency would contradict the Lardil analysis given above.

The idea that output representations contain covert structure that can influence surface structure became a basis for yet another approach to opacity, the theory of Turbidity (see Goldrick 2000 and McCarthy 2007 for discussion). Proponents of Turbidity recognize two types of relationship between segments and prosodic structure: Projection, whereby segments project as prosodic structure, and Pronunciation, whereby prosodic structure is pronounced as segments. Mismatch between the two results in opacity effects. This can be illustrated by an example from Luganda, where compensatory lengthening applies whenever a vowel is deleted to avoid hiatus. In this case, the optimal output incurs an unmotivated violation of faithfulness to the underlying vowel length, i.e. lengthening overapplies.

(3-21) *Compensatory lengthening in Luganda* (adapted from Goldrick 2000:2)

/ka + oto/ → ko:to ‘fireplace (dim.)’



On the scheme above, up-pointing arrow stands for projection relationship, down-pointing arrow stands for pronunciation relationship, while straight line indicates that there is a match between the two, i.e. that mora projected by the segment is pronounced as that segment.

According to Goldrick 2000, output candidates are evaluated by the constraints that refer to Pronounce and Project relationships. The constraint on the Realization of structure, Reciprocity$^{X}{}_{Y}$, demands that Pronounce and Project agree. When high-ranked, Reciprocity$^{X}{}_{Y}$ selects transparent candidates. In case of Luganda compensatory lengthening, Reciprocity$^{X}{}_{Y}$ is dominated by Structural Harmony constraints, Pronounce-μ demanding that all moras are pronounced and *VV prohibiting pronunciation of adjacent vowels. The opaque candidate in () above satisfies both these constraints, and is, therefore, rendered optimal.

Recall the above-mentioned case of Bedouin Arabic, where epenthesis counterfeeds vowel raising in open syllables. According to McCarthy (2007:32), in the light of Turbidity, /gabr/ → [gabur] mapping is due to the fact that [b] projects as the coda of the syllable [gab], and is pronounced as the onset of the syllable [bur].

According to McCarthy (2007:33), Turbidity is not unlike OT-based derivational approaches to opacity in that it posits an additional level of representation, distinct from both the input and the output, and because of this similarity, it is also expected to share some limitations of those approaches. In the following sections we will see what those limitations are.

According to McCarthy & Prince (1995), one of the consequences of Containment is that epenthetic segments lack featural specification and therefore cannot participate in phonological processes. However, numerous counterexamples to this claim have been found (see McCarthy & Prince 1995, Kager 1999a, McCarthy 2007 and references therein for examples and more detailed discussion). This and other considerations prompted McCarthy & Prince 1995 to abandon Containment in favour of Correspondence, which does not require the input to be contained in every output form. As we have already seen in Section 3.2, OT assuming Correspondence relation between the input and the output is not successful in dealing with opaque generalizations. Even if we abstract away from the typological inadequateness, the Containment Theory cannot be considered to be the best of possible worlds, because the increase of representational complexity and expansion of Con also means a sharp increase in computational complexity, which has unwelcome consequences for acquisition.

3.3.1.2 Local Constraint Conjunction

The approach we will consider in this section attempts to accommodate opacity in OT by means of enriching the constraint component Con. It has been proposed that the constraint component Con allows simple constraints to be joined together to form more complex constraints ("banning the worst of the worst", Prince & Smolensky 1993). This approach is called Local Constraint Conjunction (LC). Moreton & Smolensky 2002 demonstrate that LC can accomodate underapplication opacity (but, essentially, not overapplication opacity). Because of its limited scope, LC is often used as complementary to Sympathy Theory[3], about which we will talk a bit later.

(3-22) *Local Constraint Conjunction* (From Moreton & Smolensky (2002:1))

> If C_1 and C_2 are constraints, and D is a representational domain type (e.g. a segment, cluster, syllable. stem), then $(C_1 \& C_2)_D$, *the local conjunction of C_1 and C_2 in D*, is a constraint which is violated whenever there is a domain of type D in which both C_1 and C_2 are violated.

According to McCarthy (2007), the /gabr/ → [gabur] mapping of Bedouin Arabic can be accounted for in the light of this approach if we assume a high-ranking conjoined constraint $(\text{Ident(low)\&Dep})_{\text{Adj-}\sigma}$ that assigns violations to output candidates that have a raised vowel and epenthesis in adjacent syllables. Being appropriately ranked, this constraint prefers the opaque candidate [gabur] to its transparent competitor *[gibur].

According to McCarthy (2007:35), the principled problem with the local conjunction approach is that conjoined constraints refer to the proximity of processes, rather than their interaction. He provides an example from Bedouin Arabic where vowel raising applies transparently when it is preceded by the epenthetic vowel. McCarthy notes, that in this case $(\text{Ident(low)\&Dep})_{\text{Adj-}\sigma}$ cannot make the correct choice between the candidates, since it can only refer to the domain and proximity of processes, but not to their linear order.

Local Constraint Conjunction also has significant consequences for acquisition. First, it is important to determine the status of the conjoined constraints: are they innate and hence

---

[3] See McCarthy 1999 for the discussion on this point.

universal or are they constructed and hence language-specific? Both options have their pitfalls.

According to Kager (1999a:400), if conjunction of any number of any type of constraints is in principle allowed, the number of constraints in Con becomes infinite. Even if certain restrictions on the maximal number of constraints that can be conjoined and the nature of conjoinable constraints are introduced, the assumption that conjoined constraints are innate leads to a considerable increase in the computational complexity.

An alternative assumption, originally due to Fukazawa and Miglio 1998 (but see also Jesney 2005, Bonilha 2002), is that conjoined constraints are constructed by language learners when the learning data are otherwise contradictory. Thus, when the learner realizes that no permutations of markedness and faithfulness can get the opaque candidate to win, the conjoined constraint is constructed to save the day. The immediate problem with this approach is that it cannot deal with spontaneous opacity. Spontaneous opacity arises when the learner is trying to acquire transparent generalizations, so the contradictions in the learning data necessary to trigger the postulation of conjoined constraints never occur.

As noted by Idsardi 2000, another problem with the language-specific constraint conjunction is that the learner gets a mechanism alternative to constraint reranking. Recall from Chapter 2 that at a certain point in morphophonemic learning the learner has to detect that his hypothesized underlying representation is inconsistent with his current phonotactic ranking. Given the tool for conjoining constraints, the learner can just conjoin two constraints instead of discarding the faulty underlying representation. If we also assume that in addition to LC the learner needs Sympathy to cope with counterbleeding, the learning process becomes quite messy.

3.3.2 Multistratal Variants of OT

According to Bermúdez-Otero (forthcoming), SPE approaches to opacity rest on the assumption that any phonological process that seems opaque on the surface must have applied transparently somewhere else in the grammar. Bermúdez-Otero (forthcoming) restates this insight in the following way:

(3-23) *Ultimate Transparency* (from Bermúdez-Otero forthcoming:7)

> A surface phonological representation *s* is opaque in respect of a phonological generalization ***p*** if ***p*** is not true in *s*, but in some other representation *r*; *s* requires its opaque properties by virtue of its relationship with *r*.

As we have already seen, in rule-based approaches such representation *r* is an intermediate derivational stage between the input and the output. However, according to Prince & Smolensky (1993), in 'classic' parallel Optimality Theory Input → Output map has no internal structure. Therefore, OT in its original formulation cannot refer to any intermediate stages. Some scholars have recognized that precisely this property prevents OT from successfully dealing with opacity effects.

There have been several attempts to incorporate serialism in OT, which relied on the assumption that the output of one OT grammar is subject to evaluation by another OT grammar. In the Stratal OT model (see Bermúdez-Otero 2003, Bermúdez-Otero forthcoming, McCarthy 2007), output candidates pass through several different grammars ('strata'), while in Harmonic Serialism, output candidates repeatedly pass through the same grammar until there is convergence (see Prince & Smolensky 1993, McCarthy 2000, McCarthy 2007).

3.3.2.1 Harmonic Serialism

Harmonic Serialism has been proposed as the alternative to the parallel evaluation in the earliest works on OT (see Prince & Smolensky 1993). Its close semblance to rule-based serialism holds some promise that it might be more successful in dealing with opacity effects than parallel OT is. In this section, following McCarthy 2000, we will see whether Harmonic Serialism produces satisfactory results when applied to the opacity problem.

According to McCarthy 2000, in Harmonic Serialism Gen takes the input and derives from it the set of output candidates, each minimally different from the input[4]. These output

---

[4] Following McCarthy 2000, here 'minimal difference' means that only one segment at a time is affected. However, see a detailed discussion about gradualness in OT-CC in Chapter 4.

candidates are then evaluated by the grammar.The optimal output determined by the grammar is returned to Gen as the new input. Gen constructs another set of minimally divergent output candidates, and the evaluation repeats. The process continues until the optimal output of the grammar equals the input.

Recall the case of Yokuts, where the interaction of long vowel lowering and closed syllable shortening gives rise to a non-surface-apparent generalization. According to McCarthy (1999), in Yokuts mappings like /ʔili:+l/ →[ʔilel], *[ʔilil] are due to the fact that lowering is counterbled by shortening. Let's try to model this situation in the light of Harmonic Serialism.

First, Gen takes the underlying form /ʔili:+l/ and constructs a set of output candidates, such that each candidate adds only one violation of faithfulness relative to the faithful parse of the underlying representation. In Tableau (3-24) below, candidate (a) is a faithful parse, candidate (b) minimally differs from (a) in that it violates Ident(high), and candidate (c) minimally differs from (a) in that it violates Max-μ. Note that the opaque form [ʔilel] is not even in the candidate set during the first pass through the grammar, because it differs from the underlying form /ʔili:l/ in two ways: it violates both Ident(high) and Max-μ.

(3-24) *First pass*

/ʔili:+l/	No-Long-High	No-Long-Closed	Max- μ	Ident(high)
a. [ʔili:l]	*!	*		
b. [ʔile:l]		*!		*
☞c. [ʔilil]			*	

Faithful candidate (a) incurs a violation of the high-ranked markedness constraint No-Long-High. Candidate (b) fatally violates No-Long-Closed. Consequently, the unmarked candidate (c) is selected as optimal after the first pass through the grammar. Now Gen takes candidate (c) as a new input, and constructs another set of minimally different candidates.

In Tableau (3-25) below, candidate (a) is a faithful parse, candidate (b) minimally differs from the input because it adds a mora, and opaque candidate (c) minimally differs from the input in that it violates Ident(high).

(3-25) *Second pass*

/ʔilil/	No-Long-High	No-Long-Closed	Max-μ	Ident(high)	Dep-μ
💣a. [ʔilil]					
b. [ʔili:l]	*!	*			*
☹c. [ʔilel]				*!	

In (3-25), candidate (b) crucially violates both high-ranked markedness constraints and therefore is eliminated. Candidate (c), opaque the intended winner, violates Ident(high) and therefore loses to the fully faithful transparent candidate (a). Note also that both (b) and (c) are harmonically bounded by (a). Since the output of the second pass through the grammar equals the input, evaluation converges on the incorrect output [ʔilil]. Unfaithful mapping /ʔili:+l/ → [ʔilil] is enough to satisfy both No-Long-High and No-Long-Closed, so the violation Ident(high) in the unfaithful mapping /ʔilil/ →[ʔilel] is unmotivated. The situation with counterfeeding opacity is quite similar and equally problematic. The only difference is that in case of underapplication, the fatal violation incurred by the opaque candidate is the one of the high-ranked markedness constraint.

We have seen that despite its similarity to rule-based approaches, Harmonic Serialism is not a successful theory of opacity. In the next section we will consider another multistratal version of OT, which, unlike Harmonic Serialism, is based on the assumption that different strata are associated with different constraint hierarchies.

3.3.2.2 Stratal OT

According to Bermúdez-Otero (forthcoming), the insight behind Stratal OT is that there are intermediate levels of representation between underlying representation and surface representation. Each level of representation is associated with a certain morphological structure. The grammar of a language consists of several strata, each being an OT grammar characterized by a certain constraint hierarchy. Different levels of representation belong to different strata. Just like in rule-based models, the output of one stratum is the input of the next. Bermúdez-Otero 2003 identifies three core principles that Stratal OT rests upon:

(3-26) *Cyclic application* (from Bermúdez-Otero 2003:2)

> Given a linguistic expression *e* with a phonological input representation *I*, the phonological function *P* applies recursively from the inside out within a nested hierarchy of phonological domains associated with (but not necessarily fully isomorphic with) the morphosyntactic constituent structure of *e*:
>
> i.e. if I=[[x][[y]z]], then P(I)=P(P(x), P(P(y),z)).

(3-27) *Level segregation* (from Bermúdez-Otero 2003:2)

> The phonology of a language does not consist of a single function *P*, but of a set of distinct functions or 'cophonologies' $\{P_1, P_2, \ldots, P_n\}$, such that the specific function P_i applying to domains of type δ_i is determined by the type of morphosyntactic construction associated with δ_i (e.g. a stem, word or phrase).

(3-28) *Cycle-internal transparency* (from Bermúdez-Otero 2003:2)

> Each cycle involves a single pass through *Gen* and *Eval*:
>
> i.e. $P_i(\delta_i) = \mathit{Eval}_i(\mathit{Gen}(\delta_i))$

In Stratal OT, opacity is said to arise from the interaction of ordered strata. Recall the case of counterbleeding interaction of Yokuts closed-syllable shortening and long high vowel lowering in /ʔili:+l/ → $_{\text{lowering}}$ [ʔile:l] → $_{\text{shortening}}$ [ʔilel]. In rule-based phonology, the counterbleeding order means that the lowering rule is ordered before shortening rule. Stratal OT implementation is somewhat similar, the difference being that phonological generalizations are enforced by means of constraint hierarchies. Thus, in Stratal OT, Yokuts counterbleeding arises because the lowering process is enforced in Stratum_1 that is ordered before Stratum_2 where shortening occurs. This is illustrated below:

(3-29) *Stratum_1: long vowel lowering*

/ʔili:+l/	No-Long-High	Max-μ	No-Long-Closed	Ident(high)
a. [ʔili:l]	*!		*	
☞b. [ʔile:l]			*	*
c. [ʔilil]		*!		
d. [ʔilel]		*!		*

In Tableau (3-29) above, the faithful candidate (a) is ruled out by high-ranked markedness constraint against long high vowels. Unmarked candidates (c) and (d) both incur fatal violations of the faithfulness constraint Max-μ. Consequently, candidate (b), where the offending long high vowel has been lowered, is selected as optimal. However, candidate (b) is not an actual output of the grammar. Rather, it is the input into the next stratum. The crucial difference between Stratum$_1$ and Stratum$_2$ is that in the latter the markedness constraint No-Long-Closed is promoted to the position where it dominates the faithfulness constraint Max-μ.

(3-30) *Stratum$_2$: closed syllable shortening*

/ʔile:l/	No-Long-High	No-Long-Closed	Max-μ	Ident(high)
a. [ʔile:l]		*!		
☞b. [ʔilel]			*	
c. [ʔili:l]	*!	*		*
d. [ʔilil]			*	*!

In Tableau (3-30), the faithful candidate (a) is eliminated due to the violation of now-high-ranked markedness constraint No-Long-Closed. Candidate (c) violates both markedness constraints. Unmarked candidates (b) and (d) tie on the faithfulness constraint Max-μ, but (d) incurs a fatal violation of Ident(high). Therefore, the opaque form in (b) is correctly selected as the optimal output.

Despite being successful in dealing with most types of opacity, Stratal OT approach is not unproblematic. Thus, some concerns have been voiced in connection to the assignment of morphosyntactic constructions to specific strata. For example, Kager (1999a:385) notes, that in case no motivation for such assignment other than the desire to capture opaque generalizations could be found, Stratal OT would turn out to be but a variant of a serial theory employing excessively complex machinery.

As it was mentioned above, Stratal OT recognizes three levels of representation: a stem, word and phrase (see Bermúdez-Otero: forthcoming for motivation behind this subcategorization). Level assignment of phrasal categories is straightforward: they universally belong to the phrase level. Therefore, the phonological processes whose domain of application is a phrase should be enforced by the constraint hierarchy of the phrase-level co-phonology.

According to Bermúdez-Otero (forthcoming), level assignment of morphological units smaller than the word is relatively free. Importantly, however, a phonological domain created by the affixation of a root, defined as an uninflected lexical item that lacks lexical category membership, must be stem-level. In their turn, phonological domains created by the affixation of a stem may be idiosyncratically assigned to the stem or to the word level. These generalizations are summarized in Table (3-31) below:

(3-31) *Morphological operations and phonological levels* (from Bermúdez-Otero forthcoming: 74)

Morphological operation		**Phonological properties**
root-based	root-to-stem	must be stem-level
stem-based	stem-to-stem	may be stem-level or word-level
	stem-to-word	every MWord defines a word-level domain
word-based	word-to-word	

Thus, if the base of affixation is a stem, there is no other way to assign a morphological construction to a certain phonological level than by observing the phonological behavior of such a construction. This problem is known as 'co-phonology arbitrariness' (the term due to Bermúdez-Otero 1999:75). Certain languages (see Bermúdez-Otero 1999 and references therein for examples) demonstrate narrow correspondence between morphological and phonological properties of specific constructions, e.g. in KiRundi all derivational suffixes belong to the stem level, while all inflectional prefixes belong to the word-level. However, as Bermúdez-Otero (1999:75) himself notes, such behavior is "by no means the rule". Whether reliable universal non-phonological diagnostics of the level-assignment exist is still an open question.

In addition, it has been observed that Stratal OT is not enough to deal with all attested cases of opacity. To illustrate some complications the theory runs into, McCarthy (2007:41) provides the example from Bedouin Arabic, whereby deletion of high vowels stands in counterfeeding relation to raising of low vowels.

(3-32) *Counterfeeding order in Bedouin Arabic* (adapted from McCarthy 2007)

Underlying	/dafaʕ/
Deletion	N/A
Raising	[difaʕ]
Surface	[difaʕ]

According to McCarthy (2007), in Bedouin Arabic deletion of high vowels in open syllables also applies across the word boundary, resulting in mappings like /ka:tib al-ʒawa:b/ → [ka:t.bal.ʒu.wa:b] ‘writing the letter’ (see McCarthy 2007:12). It means that deletion of high vowels must be enforced by the phrase-level co-phonology, associated with the outermost stratum. Raising, in its turn, never applies across the word boundary, therefore it must be enforced by the co-phonology of some earlier stratum. However, this puts the processes in the feeding order, predicting incorrect mapping /dafaʕ/ → [dfaʕ]. This analysis suggests that Stratal OT might not be enough to deal with the full range of opaque generalizations (see McCarthy 2007 and references therein for more criticism).

Certain concerns have also been raised by the fact that Stratal OT does not have a formal mechanism to restrict the differences between the constraint rankings of different co-phonologies. Restrictions proposed so far are not uncontroversial (see McCarthy 2007 for detailed discussion and references). If unrestricted, different co-phonologies within a language can in principle vary as much as OT grammars of different languages. However, according to Kager (1999a:385), in Stratal OT analyses of different phonological phenomena rankings differ only minimally. The absence of a principled way to restrict constraint rankings can significantly increase computational complexity, which will have unwelcome consequences for acquisition.

Finally, certain scholars (e.g. Kager 1999a:385) have raised doubts about the learnability of Stratal OT grammars. Stratal OT is not straightforwardly compatible with the constraint-demotion algorithms we have discussed in Chapter 2. Computational complexity grows in proportion to the number of constraint hierarchies to be acquired. The proponents of Stratal OT, however (see Bermúdez-Otero forthcoming, Bermúdez-Otero 2003), have argued that Stratal OT can be efficiently learned with a minimum of additional machinery. In following chapters we will consider the problem of acquisition of Stratal OT grammars in somewhat more detail.

3.3.3 Horizontal Correspondence

In this section we will discuss approaches to opacity that implement the Ultimate Transparency principle and at the same time adhere to the assumption that Input → Output mapping in OT has no internal structure. To refer to such approaches, Bermúdez-Otero (forthcoming:8) uses the term 'Horizontal Correspondence'. The main insight behind the theories based on Horizontal Correspondence is that the opaque phonological representation gets its opaque properties by virtue of being related to some other output form. In Output-Output Correspondence theory, such correspondence relationship is established between the opaque form and the compositionally and semantically related transparent actual output form. In Sympathy Theory, the correspondence holds between the opaque form and a failed rival output candidate. Below we will consider relative merits and drawbacks of these two approaches.

3.3.3.1 OO-correspondence

The key assumption of the theory of output-output correspondence (see Kager 1999a, Kager 1999b) is that there is a special type of faithfulness constraints that can refer to a form that is distinct from both the input and the output. According to Kager (1999a:386), OO-correspondence can be successfully used to analyze opaque generalizations, provided that there is an output form ('base') morphologically and semantically related to the opaque output and transparent with respect to the generalization in question. In addition, according to Kager (1999b:7), the base is a free from, such that "the meaning of the affixed form contain[s] all grammatical features of its base". OO-faithfulness constraints used in the evaluation assign violation marks for every disparity between the output candidate and the base.

The chief problem with the approach is that the transparent compositionally related base is not always readily available for any opaque output form. As McCarthy (2007:45) notes, in case of /gabr/ →[gabur] counterfeeding, there is no word that would be morphologically more basic than [gabur], 'grave'. The word [gabri] 'my grave' is transparent with respect to the lack of vowel raising, but it cannot be used as a base because it is derived from [gabur], and not vice versa. Additionally, according to Kager (1999a:386), in some cases the paradigm lacks the transparent form altogether.

3.3.3.2 Sympathy Theory

In Sympathy Theory (see McCarthy 1999), the third form that participates in evaluation alongside the input and the output is a failed output candidate, or the ‘sympathetic candidate’. Sympathetic candidate is the most harmonic of the forms that obey a designated faithfulness constraint, or the ‘selector’. Another kind of faithfulness constraint, ‘sympathy constraint’, enforces identity between the sympathetic candidate and the actual output. This approach is illustrated in Tableau (3-33) below, where ☆ is used to mark a selector constraint, and ❁ is used to mark a sympathy constraint and a sympathetic candidate.

(3-33) *Counterfeeding in Sympathy Theory*

	/gabr/	*Comp-Coda	❁Id(low)	*aCV	☆Dep	Id(low)
opaque	☞a. gabur			*	*	
sympathetic & faithful	❁b. gabr	*!				
transparent	c. gibur		*!		*	*

In Tableau (3-33), Dep is a designated selector constraint. Candidate (b) obeys Dep, and therefore it is selected as a sympathetic candidate. High-ranked sympathy constraint demands that the optimal output has the same vowel quality as the sympathetic candidate. Transparent form (c) incurs a fatal violation of the sympathy constraint. Therefore, the opaque candidate (a) is correctly chosen as the winner.

Many scholars have raised serious concerns about the consequences of Sympathy Theory for acquisition (see Kager 1999a, Idsardi 2000, Bermúdez-Otero 2003). McCarthy (1999:9) admits that the learnability of Sympathy Theory is an important empirical and theoretical issue, and that it requires in-depth investigation. Inter alia, McCarthy (1999:ibid) notes that in order to account for the acquisition of sympathy effects, it is probably necessary to extend the learning algorithm proposed by Tesar & Smolensky 1993.

However, a number of works (see Dinnsen *et al*. 2000, Dinnsen 2008) use Sympathy Theory for the analysis of cases of spontaneous counterbleeding opacity. Dinnsen *et al.* 2000 consider a case of counterbleeding interaction of final obstruent deletion and vowel lengthening before voiced obstruents, which results in opaque mappings like /kæb/ → [kæ:], where vowel lengthening seems to have overapplied.

(3-34) *Some relevant constraints and a preliminary ranking* (from Dinnsen et al. 2000:327)

a. Markedness constraints

Lengthen: Avoid short vowels before voiced consonants; avoid long vowels elsewhere.

NoCoda: Avoid obstruents in codas.

b. Faithfulness constraints

ID[weight] : The length (or weight) of corresponding vowels in the in? put and output should be identical.

Max: Every input segment has a corresponding output segment (no deletion).

Ranking: NoCoda, Lengthen >> ID[weight], Max

(3-35) *Sympathy constraint*

❀**SYM**: The length (or weight) of corresponding vowels in the flower candidate and an output candidate should be identical.

(3-36) *Sympathy account of spontaneous counterbleeding* (from Dinnsen et al. 2000:334)

‘cab’/kæb/	NoCoda	❀SYM	Lengthen	✩Max	ID[weight]
a. kæb	*!	*	*		
b. ❀kæ:b	*!				*
c. kæ		*!		*	
d. ☞kæ:			*	*	*

In Tableau (3-36) above, both candidate (a) and candidate (b) satisfy selector constraint Max. However, (b) is more harmonic than (a) relative to the constraint hierarchy, therefore (b) is selected as a sympathetic candidate. High-ranked sympathy constraint demands that the optimal output has the same vowel length as the sympathetic candidate. Transparent form in (c) does not satisfy this requirement. Therefore, the opaque form in (d) is correctly selected as optimal.

Dinnsen *et al.* 2000 base their analysis on the assumption that sympathy constraints are innate, and therefore present at the initial state. Furthermore, they propose that the default ranking of sympathy constraints is such that they dominate IO faithfulness but are dominated by markedness. Importantly, such default ranking predicts the transparent form [kæ] to be optimal. In order for ❀SYM constraint to be able to choose the opaque candidate it should be crucially ranked above the worst violation of the intended winner, i.e. in this case, above Lengthen. This state of affairs might be problematic. Given the default ranking of constraints, ❀SYM can come to dominate Lengthen only if Lengthen is demoted below it in the course of acquisition. Recall from the previous chapter that the only reason the learners have for constraint demotion is so called ‘positive evidence’, i.e. the observation that some markedness constraint is violated by the actual output form of the target language. The markedness constraint Lengthen is satisfied by all output forms of English, because short vowels do not occur before voiceless obstruents, and long vowels do not occur elsewhere. Dinnsen (2008:167) suggests that the demotion of Lengthen below ❀SYM is motivated by the fact that “the child recognized some property of the target word that he had not recognized previously and that was not present in the prior transparent winner”. In other words, the ‘discovery’ of vowel length would motivate the child to demote Lengthen just below ❀SYM. In fact, in order to derive the optimal output [kæ:b], the child has to demote NoCoda below Max. Such demotion is minimal and gradual in accordance with Minimal GLA and BCD, but at no point will it result in the opaque output [kæ:]. Even if we assume that promotion of ❀SYM can in fact be motivated by positive evidence, we predict that ‘opaque’ output is an obligatory intermediate stage in the acquisition of transparent interaction of obstruent voicing and vowel lengthening in English. Scrupulous longitudinal studies are needed to check this prediction.

Another learnability problem associated with Sympathy Theory has been dubbed ‘bouquet problem’(see Bermúdez-Otero 2003, Dinnsen et al. 2000). Bouquet problem stems from the assumption that any IO-faithfulness constraint can serve to select a ❀-candidate. In the worst case scenario, then, the learner has a large choice of potential ❀-candidates for each output. McCarthy 1999 does not provide any discussion as to on what grounds the selector constraint should be chosen, and simply stipulates it. Obviously, a selector constraint cannot be externally stipulated for every case of spontaneous opacity. Dinnsen *et al.* 2000 suggest that the Elsewhere Condition can be employed when potential flower-candidates stand in subset relation, to the effect that the most inclusive candidate is selected. However, Dinnsen *et al.*

2000 note that further research is necessary to determine whether this selection principle is capable of accounting for all cases of sympathy. They also note that if there is a case where potential candidates are not in subset relation, some additional machinery might be needed to eliminate some flower-candidates.

Recall from the previous section that Sympathy Theory is not a general theory of opacity because it cannot deal with counterfeeding-on-focus. In practice it means that in order to acquire a full set of opaque effects of their language, the learners need to employ some additional machinery apart from sympathy, i.e. local conjunction. Such dichotomy, no doubt, significantly complicates the acquisition process. It also goes without saying that it is not even remotely compatible with CD algorithms discussed in Chapter 2. It is, of course, up to extensive empirical research to determine whether LC paired up with Sympathy Theory can adequately describe the full range of opacity effects that hold in developing and stable grammars. The above discussion, however, leads us to conclude that Sympathy Theory coupled with LC is not a particularly elegant theory of opacity.

3.4 Summary

In this chapter we have introduced the notion of phonological opacity and demonstrated why this phenomenon posits a principled problem for Optimality Theory in its classical formulation. We have considered different OT-compatible approaches to opacity, aimed at facilitating the analysis of misapplication at the same time preserving the advantages of constraint-based theory of grammar. The approaches to opacity we have looked at can be roughly subdivided into two large groups: those that rely on expanding some basic assumptions about the nature of underlying representations or the constraint component Con and those that make reference to forms distinct from the input and the actual output. We have seen that the latter approaches prove to be more successful at least from the typological prospective. However, even those of them that make correct predictions about the typology of opaque generalizations face a lot of difficulties with respect to the acquisition challenge. In Chapter 4 we will present a novel approach to phonological opacity called Optimality Theory with Candidate Chains, proposed by McCarthy (2007). We will consider its merits, drawbacks and principled distinction from the previous approaches, and see if it is able to account for the full range of opaque generalizations.

Chapter 4: Optimality Theory with Candidate Chains

4.1 Introduction

In this chapter we will present an extension of Optimality Theory aimed primarily at facilitating analyses of opaque generalizations. Optimality Theory with Candidate Chains, or OT-CC, was proposed by John McCarthy (2007) and represents a synthesis of OT with derivations. Below we will see, however, that though the idea itself is by no means new (see, for example, Chapter 3 for the discussion on Stratal OT and Harmonic Serialism), OT-CC is crucially different from previous derivational approaches to opacity. We will demonstrate that OT-CC successfully deals with both counterbleeding and counterfeeding opacity.

This chapter is organized as follows: in Section 4.2 we will introduce the core elements of OT-CC and briefly examine its similarity with the previous approaches to opacity; in Section 4.3 we will analyse several cases of counterbleeding and counterfeeding opacity in the light of OT-CC, and in Section 4.4 we will summarize our discussion.

4.2 Optimality Theory with Candidate Chains

Similar to the approaches considered in the previous chapter, OT-CC also relies on the Ultimate Transparency Principle (see Bermúdez-Otero: forthcoming). OT-CC differs from classic OT primarily in its understanding of what constitutes an output candidate. According to McCarthy (2007), in OT-CC a candidate is a chain of forms, rather than a single form. Candidate chains are subject to a number of restrictions. First, chains should be faithfully initiated, meaning that the first member of every chain is a fully faithful parse of the input form. In case there are several fully faithful candidates, the most harmonic of them relative to the constraint hierarchy is selected as a first form of the chain. Second, forms in the chain should accumulate differences from the input gradually. That is, every successive candidate should add a single violation of a basic faithfulness constraint in a specific location in a form (localized unfaithful mapping or LUM). Finally, the candidate chains should be harmonically improving, meaning that every successive form in the chain should be more harmonic than the preceding one relative to the given constraint hierarchy. In other words, every successive form in the chain should fare better on markedness constraints than its predecessor. Somewhat more formal wording of these requirements is given below:

(4-1) *Definition: Candidate Chain* (from McCarthy 2007:62)

A candidate chain associated with an input /in/ in a language with constraint hierarchy H is an ordered *n*-tuple of forms C = $<f_0, f_1, \ldots, f_n>$ that meets the following conditions:

a. Initial form: f_0 is the faithful parse of /in/ that is most harmonic according to H.
b. Gradualness: In every pair of immediately successive forms in C, $<\ldots, f_i, f_{i+1}, \ldots>$ $(0 \leq i < n)$, f_{i+1} has all of f_i's localized unfaithful mappings relative to /in/, plus one more.
c. Local optimality (harmonic improvement + best violation): For every pair of immediately successive forms in C, $<\ldots, f_i, f_{i+1}, \ldots>$ $(0 \leq i < n)$, where F is the basic faithfulness constraint violated by the LUM that distinguishes f_{i+1} from f_i, f_{i+1} is more harmonic according to H than f_i and every other form that differs from f_i by a different F-violating LUM.

In classic OT (see Tesar & Smolensky 1993:3), the number of output candidates generated by the function Gen is infinite. According to McCarthy (2007:65), it is a result of structure-building operations, like epenthesis, being unrestricted in Gen. According to McCarthy (2007), restrictions imposed on candidate chains in OT-CC require recursive two-way communication between Gen and Eval, since every successive link in the chain should be more harmonic than the previous. This imposes restrictions on the length and number of chains, because no operation will be allowed to apply unless its application increases the harmony of the form relative to constraint hierarchy H. Harmonic improvement, in its turn, is bounded by Characterization Theorem (Moreton 2003 as quoted in McCarthy 2007:65), originally devised for classic OT grammars and directly relevant for OT-CC. The theorem states that any OT grammar is eventually idempotent, in a sense that for any input there comes a point where the application of grammar G to its own output produces no further changes (recall from the previous chapter that this property of OT grammar guaranteed convergence for analyses based on Harmonic Serialism). In OT-CC, therefore, set of output candidates is finite. Moreover, according to Becker (2006:2), the number of forms more harmonic than the input is not only finite, in practice it is also very small. Consequently, the number of candidate chains constructed for every input is small as well. It goes without saying that such state of affairs has very attractive consequences for acquisition, since it considerably decreases computational complexity.

Gradualness requirement on candidate chains deserves detailed treatment. Of course, the important question is "What constitutes a single change?". "A single violation of some faithfulness constraint" is not a sufficient answer, because certain faithfulness constraints are known to overlap in their assessment. McCarthy (2007:78) illustrates it with the case of intervocalic voicing, which results in mappings like /pata/ → [pa.da]. The form [pa.da] violates both general faithfulness constraint Ident(voice) and positional faithfulness constraint $\text{Ident}_{\text{Onset}}$(voice). Ruling out intervocalic voicing on principled grounds is, of course, highly undesirable. Therefore, McCarthy (2007:79) suggests that only violations of *basic faithfulness constraints* are relevant for Gradualness. McCarthy 2007 (with a reference to McCarthy and Prince 1995, 1999) proposes that the following faithfulness constraints should be treated as basic: Max(x) and Dep(x) constraints that prohibit the deletion and insertion of elements of type x, Ident(f) constraints that prohibit changing particular features of such elements, and a couple of others, like Linearity that prohibits changing the linear order of elements. According to McCarthy (2007:79) these faithfulness constraints are considered basic because "they cover the full range of unfaithful mappings but they do not overlap with one another".

As it was mentioned above, a chain of forms in OT-CC resembles a sequence of intermediate forms in derivational approaches like rule-based serialism and serial versions of OT. According to McCarthy(2007:68), what crucially distinguishes OT-CC from most of other derivational approaches is the harmonic improvement requirement on intermediate forms. Stratal OT, for example, cannot implement this requirement because the central idea of the theory requires the existence of several different (and sometimes contradictory) constraint hierarchies within the same language. Therefore, if some form f_1 is more harmonic than f_2 relative to the hierarchy of stratum_1, it is possible that the same form f_1 is less harmonic than f_2 relative to the hierarchy of stratum_2. Another important difference between OT-CC and Stratal OT is that in the former requires that each successive form adds only one new faithfulness violation, while in the latter each intermediate form can in principle differ from the preceding one in any number of faithfulness violations.

McCarthy (2007:69) notes that OT-CC resembles Harmonic Serialism more closely than any other variant of OT. Recall from the previous chapter that in Harmonic Serialism the modification of the input is also limited to one faithfulness violation. One important difference between the two, according to McCarthy 2007, is that in Harmonic Serialism the intermediate form is the most harmonic of all the forms that minimally deviate from the faithful parse, while in OT-CC the first step in the chain is simply more harmonic than the faithful parse. Another difference is that in Harmonic Serialism the input into each successive pass through the grammar is the output of the preceding pass. Consequently, during the second pass and all subsequent passes faithfulness constraints “lose sight” of the lexical representation. In OT-CC, every candidate chain is initiated by the faithful parse of the underlying representation, and faithfulness constraints evaluate the relationship between the original input and successive forms in the chain.

It is also important to say a couple of words about the evaluation of candidates in OT-CC. According to McCarthy 2007, for the purposes of evaluation by the grammar candidate chains are reduced to four crucial bits of information: input, output, L-set, which is the list of LUMs incurred by all forms of the chain, and LUMSeq, which is a sequentially ordered list of LUMs of each specific chain.

The sequence of LUMs is projected from the chain. For chains consisting of only a faithful parse L-set and LUMSeq are empty. For chains involving one faithfulness violation, L-set contains one LUM, while LUMSeq is empty, since no sequential ordering is possible. There are more complex cases, however, whereby L-set contains several LUMs that cannot be ordered. We will discuss such cases below.

According to McCarthy 2007, localized unfaithful mapping, or LUM, is defined not only through the violation of a specific faithfulness constraint, but also through the locus of such violation. Locus of violation is a segment in the form that is affected by the violation of a specific faithfulness constraint. In notation proposed by McCarthy (2007:95), LUM referred to as Max@3, for example, encodes the information about the third segment in the form being affected by the violation of Max constraint, i.e. deletion. In some cases, valid candidate chains constructed on the basis of some input form /in/ have identical output forms, but differ in the order of LUMs. According to McCarthy (2007:96), such situation arises every time there is no interaction among phonological processes. In OT-CC, such chains are referred to as ‘convergent chains’.

(4-2) *Chain convergence* (from McCarthy 2007:96)

A nonempty set of chains C derived from input /in/ is convergent if and only if
(i) there exists a form [out] such that, for any chain c_i Є C, c_i = <..., out >
and
(ii) there exists a set of LUMs, L, such that, for any c_i Є C, c_i is associated with a LUMSeq l_i that is an ordering of L.

According to McCarthy (2007:96), “convergent chains are collapsed into a single candidate by merging their LUMSeqs, retaining all and only the LUM precedence relations that are common to the set of convergent chains”. The list of such precedence relations in the form of pairwise orders is referred to as a ‘reduced LUMSeq’ or just LUMSeq. For non-convergent chains LUMSeq and rLUMSeq are identical. Formal definition of a candidate in OT-CC looks as follows:

(4-3) *Candidate in OT-CC* (from McCarthy 2007:97)

A candidate is an ordered 4-tuple (in, out, L-set, rLUMSeq), where
in is a linguistic form, the input;
out is a linguistic form, the output;
L-set is a set of LUMs on *in→ out*;
and rLUMSeq is a partial ordering of a subset of L-set [or a total ordering of the L-set when a chain has no convergent mates (ibid)].

According to McCarthy (ibid), *out* is evaluated by markedness constraints, *in-out* relations encoded in L-set are evaluated by faithfulness constraints, while “evaluation of rLUMSeqs is the responsibility of **Prec** constraints, which favor certain precedence relations among the constituent LUMs of an rLUMSeq and penalize others”. Following McCarthy (2007:98), Prec(A,B) (where A and B are faithfulness constraints), “demands that every B-violating LUM be preceded and not followed by an A-violating LUM in the rLUMSeq”. Prec(A,B) can assign violation marks in two cases: first, “if there is a B-violating LUM in L-set, and this LUM is not preceded in the rLUMSeq by some A-violating LUM”; second, “if there is a B-violating LUM in L-set and it precedes some A-violating LUM”. According to McCarthy (ibid), transparent chains that compete with counterbleeding chains will be disfavoured under the first condition.

McCarthy (ibid) notes, however, that the ranking of Prec constraints is not entirely free. He introduces a metaconstraint on the ranking of Prec constraints, as cited below:

(4-4) *Metaconstraint on the ranking of Prec constraints* (from McCarthy 2007:99)

B>>Prec (A,B)

In Chapter 5 we will return to the question of formal restrictions on the ranking of the Precedence constraints. But now we are fully set to see how OT-CC deals with counterfeeding and counterbleeding opacity.

4.3 Counterbleeding and counterfeeding in OT-CC

In this section we will demonstrate that OT-CC can successfully deal with both types of opacity. However, as will be shown below, additional assumptions might be necessary to accommodate zero-terminating chain-shifts.

4.3.1 Counterbleeding

In the previous chapter we have seen the example of counterbleeding interaction of closed syllable shortening and lowering in Yokuts. For the sake of convenience, the relevant data are reproduced below:

(4-5) *Yokuts vowel alternations* (from McCarthy 1999:22)

a. Vowels are shortened in closed syllables:

/pana:/	panal	cf. pana:hin	‘might arrive/arrives’
/hoyo:/	hoyol	cf. hoyo:hin	‘might name/names’

b. Long high vowels are lowered:

/ʔili:/	ʔile:hin		‘fans’
/c’uyu:/	c’uyo:hun		‘urinates’

c. Vowels shortened in accordance with (a) are still lowered:

/ʔili:/	ʔilel		‘might fan’
/c’uyu:/	c’uyol		‘might urinate’

We have also established, following Baković 2007, that transparent application of long vowel lowering is due to the markedness constraint No-Long-High and faithfulness constraint Max-μ ranked above the faithfulness constraint Ident(high). According to Baković 2007, the process whereby long vowels are shortened in closed syllables is due to the markedness constraint No-Long-Closed dominating Max-μ. These considerations prompted us to establish the following constraint hierarchy for Yokuts:

(4-6) *Yokuts constraint hierarchy*

No-Long-High, No-Long-Closed >> Max-μ >> Ident(high)

In the previous chapter we have already seen that this ranking produces incorrect results when lowering and closed-syllable shortening interact opaquely, as in example 4-5 (c) above. The intended winner, opaque candidate [ʔilel] ends up being harmonically bounded by the transparent competitor [ʔilil]. At this point classic OT reaches a deadlock, because no ranking of constraints can result in [ʔilel] beating [ʔilil].

Let's see if OT-CC can help us to save the day. First we should construct candidate chains for the input /ʔili:+l/ and their corresponding candidates.

(4-7) *Candidate chains for the input /ʔili:+l/*

a. <ʔili:l> - faithful parse
b. <ʔili:l, ʔile:l>, Qd. No-Long-High >> Ident(high)
c. <ʔili:l, ʔilil>, Qd. No-Long-Closed >> Max-μ
d. <ʔili:l, ʔile:l, ʔilel>, Qd. No-Long-High >> Ident(high) and No-Long-Closed >> Max-μ

(4-8) *Candidates for the input /ʔili:+l/ ordered as (in, out, L-set, rLUMSeq)*

a. (/ʔili:l/, ʔili:l, Ø,Ø)
b. (/ʔili:l/, ʔile:l, {Ident(high)}, Ø)
c. (/ʔili:l/, ʔilil, {Max-μ}, Ø)
d. (/ʔili:l/, ʔilel, {Ident(high), Max-μ}, < Ident(high), Max-μ>)

Now, according to McCarthy 2007, we have to determine which Prec constraint is relevant for our analysis. In case of counterbleeding opacity, it is a Prec constraint that favors the sequence of LUMs characteristic of the intended winner.

(4-9)

Prec(Ident(high), Max-μ) – every violation of Max-μ must be preceded by the violation of Ident(high) and must not be followed by the violation of Ident(high).

In accordance with the metaconstraint on ranking, Prec(Ident(high), Max-μ) should be ranked below Max-μ. For it to affect the result of evaluation, it should also be ranked above the constraint the violation of which is the worst violation of the intended winner, i.e. above Ident(high).

(4-10) *Yokuts counterbleeding in OT-CC*

/ʔili:+l/	No-Long-High	No-Long-Closed	Max-μ	Prec(Ident(high), Max-μ)	Ident(high)
a. [ʔili:l]	*	*!			
b. [ʔile:l]		*!			*
c. [ʔilil]			*	*!	
☞d. [ʔilel]			*		*

Candidates (a) and (b) in Tableau (4-10) do not violate Max-μ, so they both vacuously satisfy the precedence constraint. Transparent candidate (c) incurs a violation of Max-μ, but this violation is not preceded by the violation of Ident(high). Therefore, candidate (c) is assigned one violation mark by the precedence constraint. Opaque candidate (d) also violates Max-μ, but this violation is preceded by the violation of Ident(high). Therefore, candidate (d) satisfies the precedence constraint. Candidates (a) and (b) are eliminated by well-formedness constraints, and transparent candidate's violation of Prec is fatal. Thus, opaque candidate (d) is correctly selected as optimal. Thus we have seen that OT-CC is able to accommodate the analysis of counterbleeding opacity.

4.3.2 Counterfeeding

Now let's see if OT-CC is equally successful in dealing with counterfeeding opacity. From the previous chapter we already know that underapplication opacity can be of two types: counterfeeding on environment and counterfeeding on focus. Bedouin Arabic provides examples of both. Consider again the relevant data, reproduced from the previous chapter:

(4-11) *Phonological alternations in Bedouin Arabic* (adapted from McCarthy 2007)

a. Short high vowels are deleted from non-final open syllables

/kit**i**b-at/	[ˈkitbat]	'it (m.) was written'
/ʃar**i**b-at/	[ˈʃarbat]	'she drank'

b. Short low vowels become high in non-final open syllables

/katab/	[kitab]	'he wrote'
/kabak/	[kibak]	'cufflink'

c. Epenthesis applies to break up final consonantal clusters

/gabr/	[gabur]	‘grave’

In Bedouin Arabic, a-raising process does not apply if its context is created by epenthesis. It results in mappings like /gabr/→[gabur], *[gibur], where raising seems to have underapplied. This is the case of counterfeeding on environment, since in serialist terms the raising fails to apply because its conditioning environment, i.e. the open syllable, was introduced too late in the derivation. Following McCarthy 2007, we assume that vowel epenthesis is due to the markedness constraint against final consonant clusters *Comp-Coda being ranked above Dep. Normal application of a-raising requires that *aCV dominates Id(low). This results in the following ranking:

(4-12) *Constraint ranking for Bedouin Arabic a-raising and epenthesis*

*Comp-Coda, *aCV >> Dep, Id(low)

In classic OT, such ranking incorrectly predicts the transparent candidate *[gibur] to win, because its opaque competitor [gabur] is ruled out by high-ranked well-formedness constraint *aCV. Let’s see if OT-CC can resolve this problem. First we have to construct candidate chains and their correspondent candidates for the input /gabr/. According to McCarthy (2007:106), they look as follows:

(4-13) *Candidate chains for the input /gabr/* (from McCarthy 2007:106)

a. <gabr> - a faithful parse
b. <gabr, gabur>, Qd. *Comp-Coda >> Dep
c. <gabr, gabur, gibur>, Qd. *Comp-Coda >> Dep and *aCV >> Ident(low)

(4-14) *Candidates for the input /gabr/ ordered as (in, out, L-set, rLUMSeq)* (from Mccarthy 2007:106)

a. (/gabr/, gabr, Ø, Ø)
b. (/gabr/, gabur, {Dep}, Ø)
c. (/gabr/, gibur, {Dep, Ident(low)}, <Dep, Ident(low)>)

Now we have to determine what Prec constraint is relevant for our analysis. In case of counterfeeding opacity, this is the Prec constraint that penalizes the transparent candidate. In our case, according to McCarthy (2007:106), it is Prec (Ident(low), Dep).

(4-15)

Prec (Ident(low), Dep) – every violation of Dep must be preceded and must not be followed by the violation of Id(low).

In accordance with the metaconstraint on the ranking of Prec, Prec(Id(low), Dep) must be ranked below Dep. In order for Prec to distinguish between the candidates, it should be ranked above the worst violation of the intended winner, i.e. above *aCV.

(4-16) *Counterfeeding-on-environment in OT-CC* (adapted from McCarthy 2007:107)

/gabr/	*Comp-Coda	Dep	Prec(Ident(low), Dep)	*aCV	Ident(low)
a. gabr	*!				
☞b. gabur		*	*	*	
c. gibur		*	**!		*

In Tableau (4-16), faithful candidate (a) vacuously satisfies the Prec constraint, because it incurs no violations of Dep. Opaque candidate (b) incurs one violation of Prec, because the violation of Dep is not preceded by the violation of Ident(low). The transparent candidate (c) incurs two violations of Prec, because the violation of Dep is followed by the violation of Ident(low). The faithful candidate (a) is eliminated by the high-ranked well-formedness constraint against complex codas. Dep cannot choose between the remaining candidates, because both (b) and (c) violate Dep once. Prec constraint prefers candidate (b) over candidate (c), because candidate (c) incurs two violations of Prec, while candidate (b) incurs only one. Thus, the opaque candidate (b) is correctly predicted to win.

OT-CC is also successful in dealing with counterfeeding-on-focus, which we will illustrate below[5]. Lubowicz 2003 (with a reference to Rubach 1984) provides example of a chain-shift in Polish. In Polish, there is a process of Nominal Strident Palatalization (NSP), whereby postalveolar voiceless fricative /ʃ/ turns into prepalatal voiceless fricative [ɕ] if it is followed by a front high vowel [i]. There is also a process called First Velar Palatalization (FVP), which turns velars /k g x/ into [tʃ ʤ ʃ] before front vocoids [i e j]. Crucially, FVP counterfeeds NSP, and therefore, unlike underlying /ʃ/, ʃ derived by FVP do not palatalize to

---

[5] Here and further we will use IPA symbols for the sake of uniformity

[ɕ]. This is the case of counterfeeding-on-focus, since NSP does not apply because its target appears too late in the derivation.

(4-17) *NSP and FVP in Polish* (from Lubowicz 2003:315)

a. Nominal Strident Palatalization: /ʃ/ → [ɕ]/_i

nom. sg.	aug.	dimin.	
gro[ʃ]	gro[ɕ]+isk+o	gro[ɕ]+ik	‘a penny’
kapelu[ʃ]	kapelu[ɕ]+isk+o	kapelu[ɕ]+ik	‘hat’

b. First Velar Palatalization: /k g x/→ [tʃ ʤ ʃ]/_i e j

nom. sg.	aug.	dimin.	
gro[x]	gro[ʃ]+ysk+o	gro[ʃ]+ek	‘bean’
gma[x]	gma[ʃ]+ysk+o	gma[ʃ]+ek	‘building’

Following Lubowicz 2003, for the purposes of our analysis, we assume that /xi/→ [ʃi] mapping is induced by ranking markedness constraint *xi above Ident(coronal), while /ʃi/ → [ɕi] mapping is induced by ranking markedness constraint *ʃi above Ident(back). Below we provide harmonically improving candidate chains and corresponding candidates for the inputs /groʃik/ and /groxek/ respectively.

(4-18) *Candidate chains for the input /groʃik/*

a. <groʃik> - a faithful parse
b. <groʃik, groɕik> - Qd. *ʃi >> Ident(back)

(4-19) *Candidates for the input /groʃik/ ordered as (in, out, L-Set, rLUMSeq)*

a. (/groʃik/, groʃik, Ø, Ø)
☞b. (/groʃik/, groɕik, {Ident(back)}, Ø)

(4-20) *Candidate chains for the input /groxek/*

a. <groxek> -a faithful parse
b. <groxek, groʃek> - Qd. *xi >> Ident(coronal)
c. <groxek, groʃek, groɕek> - Qd. *xi >> Ident(coronal), and *ʃi >> Ident (back)

(4-21) *Candidates for the input /groxek/ ordered as (in, out, L-Set, rLUMSeq)*

a. (/groxek/, groxek, Ø, Ø)
☞b. (/groxek/, groʃek, {Ident(coronal)}, Ø)
c. (/groxek/, groɕek, {Ident(coronal), Ident(back)}, <Ident(coronal), Ident(back)>)

Now we need a Prec constraint that would prefer the underapplication candidate 4-21(b) over the transparent candidate 4-21(c). In this case, Prec (Ident(back), Ident(coronal)) can do the trick.

(4-22)

Prec(Ident(back), Ident(coronal)) – violation of Ident(coronal) must be preceded and must not be followed by the violation of Ident(back).

(4-23) *Polish counterfeeding in OT-CC*

	*xi	Ident(coronal)	Prec(Ident(back), Ident(coronal))	*ʃi	Ident(back)
1. /groxek/					
a. groxek	*!				
☞b. groʃek		*	*	*	
c. groɕek		*	**!		*
2. /groʃik/					
a. groʃik				*!	
☞b. groɕik					*

In Tableau (4-23) above, the faithful candidate 1a violates a high-ranked markedness constraint *xi and is therefore eliminated. Transparent candidate 1c and opaque candidate 1b fare equally well on the faithfulness constraint Ident(coronal), but candidate 1c incurs a double violation of Prec(Ident(back), Ident(coronal)) because LUMs in its rLUMSeq are in the reversed order. Candidate 1b, in its turn, incurs only one violation of the Prec constraint because Ident(coronal)-violating LUM s not preceded by the violation of Ident(back). Therefore, candidate 1b is correctly selected as optimal. Note that Prec constraint does not interfere with the normal application of NSP: candidate 2b vacuously satisfies the Prec constraint, and is preferred by *ʃi over the faithful candidate 2a.

4.3.3 Zero-terminating chain-shifts

Above we have seen that counterfeeding-on-focus as such is not a principled problem for OT-CC. However, one particular instance of chain-shifts cannot be dealt with by means of OT-CC without additional assumptions. We are talking about so-called “zero-terminating chain-shifts”, or the chain-shifts where the second step of the derivation involves “x→Ø” mapping.

Such is the case of counterfeeding interaction of syncope and vowel raising in Bedouin Arabic, which we have already discussed in the previous chapter.

Recall that in Bedouin Arabic, syncope only affects underlying high vowels. High vowels that are the product of low-vowel raising always surface faithfully. This results in opaque mappings like /katab/ → [kitab], *[ktab], where syncope seems to have underapplied. This is the case of counterfeeding on focus, since in serialist terms the deletion process fails to apply because the relevant segment was introduced too late in the derivation. Following McCarthy (2007:103), we have assumed that normal application of syncope requires that markedness constraint against high vowels in open syllables *iCV should dominate Max. Following McCarthy (1999:7), we have also assumed that a-raising process is due to the markedness constraint *aCV ranked above Ident(low). This results in the following ranking:

(4-24) *Constraint ranking for Bedouin Arabic a-raising and syncope*

*iCV,*aCV >> Max, Ident(low)

Ranking in (4-24) incorrectly predicts the transparent candidate *[ktab] to win, because the intended winner, opaque candidate [kitab], is ruled out by the high-ranked wellformedness constraint *iCV. Let's see if OT-CC can help us fix this problem.

First, we have to construct candidate chains for the input /katab/ and their corresponding candidates.

(4-25) *Candidate chains for the input /katab/*

a. <katab> - faithful parse
b. <katab, kitab>, Qd. *aCV >> Ident(low)
c. <katab, ktab>, Qd. *aCV >> Max

Note that **<katab, kitab, ktab> chain is invalid, because the form *ktab* does not contain the violation of Ident(low) incurred by the previous form *kitab*. According to McCarthy (2007:130), this is because of the "vacuous satisfaction property" of Ident(F) constraints, which means that constraint of the type Ident(F) is not violated if the feature it refers to does not have an output correspondent.

(4-26) *Candidates for the input /katab/ ordered as (in, out, L-set, rLUMSeq)*

a. (/katab/, katab, Ø, Ø)
b. (/katab/, kitab, {Ident(low)}, Ø)
c. (/katab, ktab, {Max}, Ø)

Now we have to choose the precedence constraint relevant for the analysis. It should be the constraint that penalizes the transparent candidate, in our case the only such constraint is Prec(Ident(low), Max), since Prec(Max, Ident(low) is vacuously satisfied by the transparent candidate).

(4-27)

Prec(Ident(low, Max)) – every violation of Max must be preceded and must not be followed by the violation of Id(low).

In accordance with the metaconstraint on the ranking of Prec, Prec(Ident(low), Max) should be dominated by Max. In order to be able to distinguish between the candidates, Prec constraint should be ranked above the worst violation of the intended winner, i.e. above *iCV. But the normal application of syncope requires that *iCV dominates Max. Therefore, every ranking of Prec that would chose the opaque candidate *kitab* would also interfere with the normal application of syncope. This is illustrated in Tableau (4-28) below:

(4-28) *Zero-terminating chain shift in OT-CC*

	*aCV	Max	Prec (Ident(low), Max)	*iCV	Ident(low)
/katab/					
a. katab	*!				
☞b. kitab				*	*
c. ktab		*!	*		
/ʃaribat/					
💣a. ʃa.ri.bat				*	
☹b. ʃar.bat		*	*		

Note that even if Prec(Ident(low), Max) is ranked above *iCV >> Max, in violation of the metaconstraint, it will still preclude the normal application of syncope in /ʃaribat/ → [ʃarbat]. Unfortunately, we have to conclude that OT-CC as such is not equipped to deal with zero-terminating chain-shifts, under the assumption that input-output faithfulness is induced by

Ident(F) constraints[6]. However, according to McCarthy (2007:130), zero-terminating chain shifts can be easily accommodated with the help of positional faithfulness to prominent elements. Constraints requiring greater faithfulness to prominent elements are not new (see Beckman 1997, Beckman 1998, Jesney 2005)[7]. According to McCarthy (2007:130), low vowels enjoy greater prominence than non-low vowels - they have greater duration and amplitude and attract stress (see also references in McCarthy 2007). Constraint Max-A ranked above *iCV will correctly prefer the raising candidate *kitab* over syncope candidate **ktab* that does not contain a segment correspondent to underlying /a/. It will not interfere with normal application of syncope, because syncope applies to underlying high-vowels only.

4.4 Summary

In this chapter we have outlined a novel theory of opacity - OT-CC, proposed by McCarthy 2007. Despite the fact that OT-CC has much in common with the previous theories of opacity, it represents a major improvement with respect to those theories, in that it is able to account for both counterbleeding and counterfeeding opacity with very little additional machinery, while at the same time preserving the parallel evaluation principle of classic OT. In Chapter 5 we will consider the status of Precedence constraints in more detail, and discuss the possibility of further improvements of OT-CC with the view to increase its descriptive adequacy.

---

[6] However, see McCarthy (2007:142) for the Prec-based analysis assuming Max(F) constraint.

[7] Note that faithfulness to prominent elements (e.g. in Beckman 1997, 1998) is required independently of opacity (see McCarthy 2007 for the discussion).

Chapter 5: Metaconstraint on Ranking

5.1 Introduction

In Chapter 4 we have already introduced the fact that ranking of Precedence constraints is subject to certain restrictions. According to McCarthy (2007:98), "[a]lthough Prec constraints [...] are ranked and violable, their ranking is not entirely free". McCarthy proceeds with saying that "though Prec(A,B) obviously *depends on* whether B is violated, [it] must never *affect* whether B is violated" and "in OT, the only way to ensure this is to require that Prec(A,B) never dominate B"(italics are original). To this end, McCarthy(2007:99) postulates the ranking meta-constraint in the form B >> Prec(A,B). As any assumed universal ranking, the metaconstraint is an extra theory-internal stipulation, and as such it is undesirable. In this chapter we will consider several cases that might constitute evidence in favour of discarding the metaconstraint on the ranking of Prec.

The chapter is organized as follows: in Section 5.2 we will present and discuss three reasons in favour of discarding the metaconstraint on the ranking of Prec; in Section 5.3 we will outline our proposal concerning the status of Prec constraints, and in Section 5.4 we will briefly summarize our discussion.

5.2 Three reasons to discard the metaconstraint

Before discussing the possibility to discard the metaconstraint on the ranking of Prec, we should probably consider the reason why it was postulated on the first place. According to McCarthy (2007), the reason for postulation of the metaconstraint comes from the analysis of counterbleeding interaction of palatalization and high-vowel syncope in Bedouin Arabic.

According to McCarthy (2007:99), in Bedouin Arabic the velar stops /k/ and /g/ are palatalized to [k^j] and [g^j] when adjacent to the front vowel [i]. McCarthy (2007:187) also provides evidence that in Bedouin Arabic short high vowels are deleted from non-final open syllables. Interaction of syncope and palatalization leads to counterbleeding opacity, where the velar palatalizes even when the high vowel conditioning the palatalization is deleted. Consider the following data:

(5-1) Palatalization and syncope in Bedouin Arabic

a. Velar palatalization in Bedouin Arabic (from McCarthy 2007:99)

[ru:g]	'be calm'	[rawwig^{j}]	'do not make noise!'
[gu:l]	'say'	[g^{j}i:l]	'it was said'

b. Syncope of /i/ in Bedouin Arabic (from McCarthy 2007:187)

/kit**i**b-at/	[ˈkitbat]	'it (m.) was written'
/ʃar**i**b-at/	[ˈʃarbat]	'she drank'

c. Opaque interaction of palatalization and syncope (from McCarthy 2007:100)

/ħa:kim-i:n/	[ħa:k^{j}mi:n]	'ruling (m. pl.)'
/kitib-t/	[k^{j}tibt]	'you (m.sg.) were written'

To deal with these phenomena, McCarthy (2007: 93-94) proposes the following constraints:

(5-2)

***iCV** (dominating Max)
Short high vowels are prohibited in open syllables

***ki** (dominating Ident(back))
Sequences of a plain velar and a front vowel are prohibited

Prec (Ident(back), Max)
Max-violating LUM should be preceded and should not be followed by Ident(back)-violating LUM

Harmonically improving candidate chains for the input /ħa:kim-i:n/ constructed on the basis of the transparent ranking of the Bedouin Arabic are given below (note, that Prec constraints do not participate in the construction of candidate chains).

(5-3) *Harmonically improving candidate chains for the input /ħa:kim-i:n/ and their LUMSeqs* (from McCarthy 2007:100)

a. <ħa:kimi:n>	<>
b. <ħa:kimi:n, ħa:k^{j}imi:n>	<Ident(back)>
c. <ħa:kimi:n, ħa:k.mi:n >	<Max>
☞d. < ħa:kimi:n, ħa:k^{j}imi:n, ħa:k^{j}.mi:n >	<Ident(back), Max>

(5-4) *Harmonically improving chains for the input /ʃarib-at/ and their LUMSeqs*

a. <ʃa.ri.bat>	<>
☞b. <ʃar.bat>	<Max>

(5-5) *Palatalization and syncope in Bedouin Arabic* (adapted from McCarthy 2007:101)

	***iCV**	***ki**	**Max**	**Prec(Ident(back), Max)**	**Ident(back)**
1. /ħa:kim-i:n/					
a. ħa:.ki.mi:n	*!	*			
b. ħa:.kʲi.mi:n	*!				*
c. ħa:k.mi:n			*	*!	
☞d. ħa:kʲ.mi:n			*		*
2. /ʃarib-at/					
a. ʃa.ri.bat	*!				
☞b. ʃar.bat			*	*	

In Tableau (5-5), candidates (1a) and (1b) are eliminated by high-ranked markedness constraint against high vowels in open syllables. Candidates (1c) and (1d) tie on the faithfulness constraint Max, but (1c) incurs a fatal violation of the Precedence constraint, because Max-violating LUM in its LUMSeq is not preceded by Ident(back)-violating LUM. Thus, opaque candidate (1d) is correctly selected as optimal. In case of transparent application of syncope, the faithful candidate in (2a) is eliminated by the high-ranked *iCV constraint. This makes (2b) the winner, regardless its violation of Prec(Ident(back), Max).

In OT, every legal permutation of constraints is supposed to yield an existing grammar. Therefore, if we discard the metaconstraint, the grammar where Prec(Ident(back), Max) dominates *iCV becomes theoretically possible. McCarthy (2007:101) stresses that ranking Prec (Ident(back), Max) above *iCV would yield a grammar where syncope is discouraged unless it counterbleeds palatalization. Such hypothetical situation is illustrated below.

(5-6) *Effect of violating the ranking meta-constraint* (adapted from McCarthy 2007:102)

/ʃarib-at/	**Prec(Ident(back),Max)**	***ki**	***iCV**	**Max**	**Ident(back)**
💣a. ʃa.ri.bat			*		
☹b. ʃar.bat	*!			*	

In Tableau (5-6), candidate (b) incurs a fatal violation of Prec(Ident(back),Max) because Max-violating LUM in its LUMSeq is not preceded by Ident(back)-violating LUM.

According to McCarthy (ibid), the state of affairs illustrated in Tableau (5-6) “goes beyond anything encountered in attested opaque phonology” and “no real phonological system does or could work this way, so we need a language-independent explanation for why this never happens”. In order to rule out the unwanted scenario McCarthy proposes to postulate meta-constraint on the ranking of Prec that would ensure that Prec (Ident(back), Max) can never dominate Max.

5.2.1 Obligatorily counterbleeding processes

Matthew Wolf (2008:350), however, argues that the scenario modelled in Tableau (5-6) is, in fact, attested. He presents a case of Chimwi:ni, where vowel shortening is blocked when there is no spirantization for it to counterbleed. Wolf calls the phenomenon of Chimwi:ni “an obligatorily counterbleeding process”.

(5-7) *Obligatorily counterbleeding processes* (from Wolf 2008:350)

> A B-violating process occurs just in case it would counterbleed an A-violating process. If no A-violating process occurs, the B-violating process does not occur.

According to Wolf (2008:350-351), in Chimwi:ni the perfective suffix /-i:łe/ triggers spirantization of certain preceding consonants. When spirantization occurs, the initial vowel of the suffix shortens.

(5-8) *Chimwi:ni spirantization* (from Wolf 2008:351)

/p, t, t̪/ → [s]
/k/ → [ʃ]
/ł/ → [z]

(5-9) *Chimwi:ni vowel shortening* (from Wolf 2008:351)

[ku-łipa]	‘to pay’	/łip-i:łe/	→	[łis-iłe]	‘he paid’
[ku-lat̪a]	‘to let go’	/lat̪-i:łe/	→	[las-iłe]	‘he let go’
[x-ʃi:ka]	‘to hold’	/ʃi:k-i:łe/	→	[ʃi:ʃiłe]	‘he held’
[x-kuła]	‘to grow’	/kuł-i:łe/	→	[kuziłe]	‘he grew’

The suffix vowel fails to shorten when it is not preceded by consonant that undergoes spirantization (from Wolf 2008:351).

(5-10)

/pamb-i:ɬe/	→	[pamb-i:ɬe]	‘he decorated’
/kos-i:ɬe/	→	[kos-e:ze]	‘he made a mistake’
/set-i:ɬe/	→	[set-e:ɬe]	‘he stamped on’

Some stems with final consonants other than [p, t, t̪, k, ɬ] exceptionally undergo spirantization, and in such cases shortening of the suffix vowel does occur (from Wolf 2008:ibid).

(5-11)

/big-i:ɬe/	→	[biš-iɬe]	‘he hit’
/law-i:ɬe/	→	[laz-iɬe]	‘he went out’

On the basis of these data, Wolf (ibid) concludes the following: first, assuming that the length of the suffix vowel triggers the spirantization of the preceding consonant, Chimwi:ni presents a case of counterbleeding interaction between spirantization and shortening, since shortening removes the context for spirantization, but spirantization occurs nevertheless; second, shortening of the perfective suffix vowel is blocked just in case there is no spirantization for it to counterbleed. Therefore, Chimwi:ni exhibits exactly the kind of scenario McCarthy 2007 wanted to rule out by means of meta-constraint on the ranking of Prec. The existence of a natural language violating the metaconstraint is the first, and, probably, the most important argument in favour of discarding the metaconstraint.

5.2.2 Non-derived environment blocking

Another reason for discarding the mataconstraint on the ranking of Prec comes from the analysis of DEEs. Below we will illustrate it with the example from Polish. In Polish (see Rubach 1984:110-111), velars /k, g, x/ change into [tʃ, ʤ, ʃ] when followed by a front vowel or a glide. In rule-base phonology, this process is called ‘First Velar Palatalization’ (see Rubach 1984:ibid).

(5-12) *First Velar Palatalization in Polish*

a. krzy/k/	‘shout’	→	krzy[ʧ]+e+ć	‘to shout’
b. słu/x/	‘hearing’	→	sły[ʃ]+e+ć	‘to hear’
c. móz/g/	‘brain’	→	móż[ʤ]+ek	‘brain (dim.)’

If thus derived [ʤ] is preceded by an obstruent (as in example 5-12 (c) above), it surfaces faithfully. If it is preceded by a sonorant, it spirantizes to [ʒ] (see Rubach 1984:119).

(5-13)

a. ro[g]	‘horn’	→	ro[ʒ]+ek	‘horn (dim.)’
b. snie[g]	‘snow’	→	snie[ʒ]+ek	‘snow (dim.)’

There are, however, a number of exceptional lexical items (loanwords), whereby spirantization process fails to apply. Consider the following (from Rubach 1984:120):

(5-14)

a. parkin[g]	‘parking’	→	parkin[ʤ]+ek	‘parking (dim.)’
b. trenin[g]	‘training’	→	trenin[ʤ]+ek	‘training (dim.)’

According to Rubach (1984: ibid), such exceptions prove that /g/ to [ʒ] mapping is the result of spirantization of /ʤ/, rather than one-step change. In his own words (Rubach 1984:ibid), it proves that “the intermediate stage /ʤ/ has a linguistic (“psychological”) reality”. Summary of Polish alternations discussed so far is given below:

(5-15)

	First Velar		Spirantization	
/g/	→	(/dʒ/)	→	[ʒ]
/k/	→	[tʃ]		
/x/	→	[ʃ]		

Note that the non-derived (underlying) instances of /ʤ/ surface faithfully in spirantization contexts (see Gussmann 2007:82):

(5-16)

bry[ʤ]-a ‘bridge (gen.pl)’ → bry[ʤ]-yst-a ‘bridge player’

Traditionally, the Polish data presented above have been analysed as the case of non-derived environment blocking, whereby spirantization of *ʤ* is blocked unless *ʤ* is derived by the prior application of First Velar (see Wolf 2008 and references therein).

In LP, certain phonological processes are prevented from application in nonderived environments by Strict Cyclicity principle (see Rubach 1984, Lubowitz 2002).

(5-17) *Strict Cycle Condition* (from Kiparsky 1982:4 as quoted in Lubowitz 2002:244)

> a. Cyclic rules apply only to derived representations.
> b. *Definition:* A representation ϕ is derived w.r.t. rule R in cycle j iff ϕ meets the structural analysis of R by virtue of a combination of morphemes introduced in cycle j or the application of a phonological rule in cycle j.

In the light of this approach, Spirantization is proclaimed to be cyclic, and is therefore restricted to the context created by the prior application of the First Velar. SCC itself is not unproblematic. To start with, according to Rubach (1984:238), there are no clear formal criteria that would allow to classify certain rules as cyclic. The proposals made to this end, as Rubach 1984 notes, are not fully tenable. In addition, it was claimed that SCC appears to be empirically false (see Bermúdez-Otero 2003 and references therein). Even if we disregard these obvious drawbacks, having two separate tools to deal with different types of underapplication seems redundant (a point also made in Wolf 2008).

According to Bermúdez-Otero (2003:3), “there is no room in Stratal OT for the Strict Cycle Condition”. Having rejected LP device for dealing with NDEB, Stratal OT as such, according to Wolf(2008:339), “cannot model DEEs which occur in phonologically-derived rather than morphologically derived environments”. To illustrate his point, Wolf 2008 considers two lexical items that are morphologically identical, and yet behave differently with respect to spirantization:

(5-18) *Palatalization and spirantization of underlying /g/*

dron[g]-ĭk-ĭ → drõw̃[ʤ]-ek → drõw̃[ʒ]-ek ‘pole-diminutive’

(5-19) *Underlying /ʤ/ surfaces faithfully*

brɨ[ʤ]-ĭk-ĭ → brɨ[ʤ]-ek ‘bridge-diminutive’

According to Wolf (2008:339), Stratal OT cannot distinguish between derived and underlying *ʤ* in examples (5-18) and (5-19), “because these words contain the exact same affixes and therefore pass through the exact same strata with the exact same rankings.”

Let’s try to analyze this case in the light of Stratal OT. For our analysis we will need the following constraints (Wolf 2008:249):

(5-20)

***KE** (dominating Ident(place))
One violation-mark for every instance of a velar followed by a front vowel

***ʤ** (dominating Ident(contin))
One violation mark for every instance of [ʤ]

As we have seen above, in Polish spirantization normally applies to the result of First Velar. It means that palatalization and spirantization processes stand in feeding order, such that palatalization precedes (and feeds) spirantization. In Stratal OT, it means that palatalization is enforced by the co-phonology of some stratum$_1$ that precedes stratum$_2$ where spirantization is enforced. Spirantization, as we know, does not apply across the word boundary. Therefore, stratum$_2$ cannot be phrase-level. It follows, then, that spirantization must be word-level, and First Velar must be stem-level. Consequently, the words /drong-ĭk-ĭ/ and /brɨ[ʤ]-ĭk-ĭ/ have the domain structure as in (5-21), assuming that the suffix /ĭk-ĭ/ is stem-level. It should be noted, that the attribution of the suffix to this particular domain is not crucial - it might as well be word-level. What matters is that /drong-ĭk-ĭ/ and /brɨ[ʤ]-ĭk-ĭ/ must have the same domain structure.

(5-21)

a. [PL[WL[SL drong + ĭk-ĭ]]] ‘pole-diminutive’
b. [PL[WL[SL brɨʤ + ĭk-ĭ]]] ‘bridge-diminutive’

So, the input into stem-level co-phonology is the form /drong + ĭk-ĭ/, and the task of the grammar at this point is to correctly implement velar palatalization process active in the Polish language. The trick can be done by ranking markedness constraint *KE, which bans velars followed by front vowels, above the faithfulness constraint Ident(place).

(5-22) *Stem-level co-phonology*

	*KE	Ident(place)	Ident(contin)	*ʤ
1./drong + ĭk-ĭ /				
a. dron[g]ek	*!			
☞b. dron[ʤ]ek		*		*
c. dron[ʒ]ek		*	*!	
2./ krok + i-ć /				
a. kro[k]ić	*!			
☞b. kro[tʃ]ić		*		
c. kro[ʃ]ić		*	*!	
3./brɨʤ + ĭk-ĭ/				
☞a. brɨʤek				*
b. brɨgek	*!			
c. brɨʒek			*!	

The output of the stem-level co-phonology is also the input of the next level. The task of the word-level co-phonology is to implement the spirantization of /ʤ/. This is achieved by ranking the markedness constraint *ʤ above the faithfulness constraint Ident(contin).

(5-23) *Word-level co-phonology*

	*KE	*ʤ	Ident(place)	Ident(contin)
1. /dronʤek/				
a. drõw̃[g]ek	*!			
b. drõw̃[ʤ]ek		*!		
☞c. drõw̃[ʒ]ek				*
2. /kro[tʃ]ić/				
a. kro[k]ić	*!			
☞b. kro[tʃ]ić				
c. kro[ʃ]ić				*!
3./brɨʤ + ĭk-ĭ/				
☹a. brɨʤek		*!		
b. brɨgek	*!			
💣c. brɨʒek				*

The problem with the above analysis stems from the fact that in Stratal OT, word-level co-phonology cannot access the input to stem-level co-phonology, or lexical representation of the output candidate. Therefore, from the point of view of Stratal OT, the underlying /ʤ/ of /brɨʤ-ĭk-ĭ/ → [brɨʤ-ek] mapping is indistinguishable from the derived /ʤ/ of /drong + ĭk-ĭ/ →/drondʒek/ → [drõw̃ʒek] mapping during the pass through the word-level co-phonology. Forms /drong + ĭk-ĭ/ and /brɨʤ-ĭk-ĭ/ have identical morphological structure and therefore they must be subject to evaluation by exactly the same co-phonologies. Stratal OT, therefore, predicts the counterfactual mapping /brɨʤ-ĭk-ĭ/ - [brɨʒ-ek]**.**

Martin Krämer (personal communication) suggested an alternative treatment of the Polish data. He proposed to reformulate the context of the spirantization process to the effect that it would apply to underlying /g/ preceded by a sonorant and followed by a front vowel or /j/ and change it into [ɣ]. First Velar would then apply to the outcome of Spirantization, palatalizing /ɣ/ to [ʒ]. In case of underlying [ʤ] Spirantization would not apply since its structural description is not met, so underlying [ʤ] would surface faithfully. This account is formulated below:

(5-24)

	Spirantization		Palatalization	
/g/	→	/ɣ/	→	[ʒ]
/k/	→	N/A	→	[tʃ]
/x/	→	N/A	→	[ʃ]

If we want to use this approach in Stratal OT analysis, we have to assume that spirantization applies at an earlier level than First Velar. Thus, if First Velar is word-level, spirantization should be stem-level. This state of affairs, though appealing for Stratal OT, is associated with a number of typological problems.

Rubach (1984:113) provides a number of examples from Polish where First Velar fails to apply in unaffixed words. For such cases, our analysis predicts voiced velar fricative [ɣ] to surface, since the structural description of the reformulated spirantization rule is met. But this is not what actually happens in Polish. Consider the following:

(5-25) *Exceptions from First Velar* (from Rubach 1984:113)

a. ewan/g/elia	'gospel'	→	ewan[g]elia
b. a/g/ent	'agent'	→	a[g]ent

In fact, in Polish there is no evidence of a process that would turn underlying /g/ to [ɣ]. In Polish, [ɣ], along with [x], is considered to be an allophone of /x/. Underlying /x/ surfaces as [ɣ] only before voiced obstruents, also across the word-boundary. Consider the following examples:

(5-26) *Occurrence of [ɣ]* (from Gussman 2007:85)

a. Bochdan	'personal name'	→	[boɣdan]
b. klechda	'folk tale'	→	[klɛɣda]
c. dach był	'the roof was'	→	[daɣbɨw]

Therefore, we can safely assume that the process whereby underlying /g/ is spirantized to [ɣ] is inactive in Polish. In addition, according to Rubach (1984), spirantization also affects instances of /ʤ/ produced by affricate palatalization, resulting in mappings like /dz/ → [ʤ] → [ʒ]. If we assume that affricate palatalization feeds spirantization, which in turn feeds First Velar, we again predict underlying /ʤ/ to spirantize. Therefore, it seems that in order to accommodate NDEB effects in Stratal OT, additional machinery similar to SCC might be needed.

Notably, OT-CC is capable of dealing with NDEB effects without any additional machinery. In case of non-derived environment blocking, a certain process (here: spirantization) is blocked unless it is preceded by some other process. This is exactly the insight expressed by Prec(A, B) constraint, which is violated whenever a process B applies without being preceded by process A (see Wolf 2008 for more discussion on this point). To illustrate it, below we will outline OT-CC analysis of Polish NDEB presented by Wolf (2008:244-246, 249-252).

In addition to the constraints introduced above, Wolf (2008:249) postulates the following precedence constraint:

(5-27)

Prec(Ident(place), Ident(contin))
One violation mark whenever a candidate contains Ident(contin)-violating LUM not preceded by Ident(place)-violating LUM, and whenever a candidate contains Ident(contin)-violating LUM followed by Ident(place)-violating LUM.

For underlying /gi/ sequences it is possible to construct the following harmonically improving candidate chains:

(5-28) *Harmonically improving candidate chains for the input /gi/* (from Wolf 2008:250)

a. <gi>
rLUMSeq: Ø

b. <gi, ʤi>
rLUMSeq: <Ident(place)>

c. <gi, ʤi, ʒi>
rLUMSeq: <Ident(place), Ident(contin)>

(5-29) *Spirantization in Polish* (from Wolf 2008:250)

/gi/	**Prec(Ident(place), Ident(contin))**	***KE**	***ʤ**	**Ident(place)**	**Ident(contin)**
a. [gi]		*!			
b. [ʤi]			*!	*	
☞c. [ʒi]				*	*

In Tableau (5-29), Prec constraint is vacuously satisfied by candidates (a) and (b), because none of them violates Ident(contin). Candidate (c) satisfied Prec constraint because the Ident(contin)-violated LUM in its LUMSeq is preceded by Ident(place)-violating LUM. Candidate (a) crucially violates markedness constraint *KE, while candidate (b) violates *ʤ. Therefore, candidate (c) is correctly selected as optimal. Therefore, OT-CC correctly predicts that /ʤ/ derived from the underlying /g/ will spirantize. Note also that Prec(Ident(place), Ident(contin)) will not interfere with the application of First Velar, because palatalization involves only the violation of Ident(place) constraint.

Let's see if the above ranking can capture the fact that underlying /ʤ/ is immune to spirantization. For underlying /ʤ/, it is possible to construct the following harmonically improving candidate chains:

(5-30) *Harmonically improving candidate chains for the input /ʤ/* (from Wolf 2008:251)

a. <ʤ>
rLUMSeq: Ø

b. <ʤ, ʒ>
rLUMSeq: <Ident(contin)>

(5-31) *NDEB in Polish* (from Wolf 2008:251)

/ʤ/	**Prec(Ident(place), Ident(contin))**	***KE**	***ʤ**	**Ident(place)**	**Ident(contin)**
☞a. [ʤ]			*		
b. [ʒ]	*!				*

In Tableau (5-31), spirantizing candidate (b) incurs a fatal violation of Prec constraint, because Ident(contin)-violating LUM in its LUMSeq is not preceded by Ident(place)-violating LUM. Faithful candidate (a), in its turn, vacuously satisfies Prec constraint and is correctly selected as optimal. The analysis, therefore, correctly predicts spirantization of underlying /ʤ/ to be blocked.

Thus, following Wolf 2008, above we have demonstrated that OT-CC provides a useful tool for analyzing NDEB effects. Wolf (2008:56) emphasizes that the OT-CC analysis has a considerable advantage over the LP treatment of nonderived environment blocking, because OT-CC analysis does not require additional machinery to deal with NDEB: all the theoretical tools used for the above analysis have been independently proposed to account for opacity effects (see McCarthy 2007). However, we can only achieve the demonstrated result if we discard the metaconstraint on ranking of Prec, proposed in McCarthy 2007.

As it was mentioned above, in order to account for palatalization and spirantization of /g/, both Ident(place) and Ident(contin) should be ranked below relevant markedness constraints, *KE and *ʤ respectively. Following the metaconstraint on ranking, Prec (Ident(place), Ident(contin)) should be obligatorily dominated by Ident(contin), and therefore, by transitivity, also by both markedness constraints. However, such ranking incorrectly predicts the underlying /ʤ/ to spirantize. The analysis following the meta-constraint on ranking is presented below:

(5-32) *Effect of observing the metaconstraint in NDEB analysis*

	***KE**	***ʤ**	**Ident(place)**	**Ident(contin)**	**Prec(Ident(place), Ident(contin))**
/gi/					
a. [gi]	*!				
b. [ʤi]		*!	*		
☞c. [ʒi]			*	*	
/ʤ/					
☹a. [ʤ]		*!			
💣b. [ʒ]				*	*

Thus, the analysis of non-derived environment blocking requires the metaconstraint on the ranking of Prec to be discarded. This can be done at no cost for the typological adequacy of the theory. Quite the contrary - as we have demonstrated above, the existence of obligatorily counterbleeding processes removes the need for the metaconstraint.

We should probably mention that the Polish example considered above is not uncontroversial. Recall that LP lacks formal mechanism to divide rules into cyclic and post-cyclic. According to Rubach (1984), spirantization effects could as well be modelled if the rule applied postcyclically. The cyclic status of spirantization, according to Rubach (1984:120) is "imposed by the theory", specifically, by the fact that spirantization must be ordered before other rules that are considered to be cyclic. Rubach continues to say that regarding spirantization as cyclic makes different predictions than regarding it to be postcyclic. Namely, according to Rubach (1984:121),

> "On the first view borrowings such as *brydż* 'bridge' and *banjo* 'banjo' cannot restructure to **bryż* and **banżo* (and indeed they do not) since /ʤ/ is "nonderived" (underlying). However, on the second view *brydż* and *banjo* would now be regarded as exceptions to Spirantization and a prediction would be made that, if nativized, they will ultimately end up as *[brɨʒ] and *[banʒo]."

Actually, even cyclic spirantization is not exceptionless in Polish. Recall from (5-14) above that spirantization also exceptionally fails to apply to derived /ʤ/ in diminutive forms of certain foreign borrowings e.g. parkin[g] - parkin[ʤ]+ek vs. pingpon[g] – pingpon[ʒ]ek. This state of affairs is equally problematic for both OT-CC and Stratal OT, since underlying forms /parking+ek/ and /pingpong+ek/ are indistinguishable.

It is possible that loanwords like *parking, brydż* and *banjo* owe their exceptional behaviour to the fact that to date they have been only partially nativized. The similar point is also made by Čavar (2005:19), who observes that /ʤ/ itself is a minor phoneme in Polish. To the best of our knowledge, in native Polish words underlying /ʤ/ never appears in the spirantizing context. Consider the following examples from Gussmann (2007):

(5-33)

/drɔʒʤ-ɨ/	→	[drɔʒʤɨ]	‘yeast’
/ʤʤɔvɲits-a/	→	[ʤʤɔvɲitsa]	‘earthworm’
/ʤʤ- ɨst-ɨ/	→	[ʤʤɨstɨ]	‘rainy’

Čavar (2005:ibid) suggests that Polish data in (5-14) and (5-16) might be accounted for if we assume that there is some faithfulness constraint ranked above markedness constraints inducing spirantization. According to Smolensky, Davidson and Jusczyk (2004:338), cases when certain segments and clusters disallowed in the recipient language surface faithfully in borrowings are not uncommon. Smolensky et al. (2004:340) attribute this exceptional behaviour to a ‘floating faithfulness constraint’, a constraint that is allowed to move within its range in the hierarchy rather than being fixed in one position. According to Smolensky et al. (2004:342), “[i]n the face of non-native inputs, speakers can, when sufficient cognitive resources are allocated, elevate a faithfulness constraint from its base position to a higher position within its floating range”. If we apply this approach to the Polish case under analysis, we can assume that the exceptional behaviour of certain loanwords is due to IdentContin constraint being elevated to a position where it dominates a markedness constraint *ʤ.

However, even if such view on Polish data is justified, it does not crucially affect the outcome of our above discussion. According to Wolf (2008:262), DEEs in phonologically-derived environments are quite common, and Polish NDEB is by no means an isolated case (see Wolf 2008 and references therein for more examples). Stratal OT as such, as we have shown above, lacks machinery necessary to deal with DEEs in phonologically-derived environments. OT-CC, in its turn, can easily accommodate DEEs of different kinds (see Wolf 2008 for the analysis of DEE in morphologically-derived environments), if the metaconstraint on the ranking of Prec is discarded.

5.2.3 Spontaneous opacity

According to Wolf (2008:353), the third reason for discarding meta-constraint on ranking comes from child acquisition data, namely, from cases of spontaneous, or non-target-like, opacity. Following Wolf 2008, below we will illustrate this claim with the example of spontaneous counterfeeding. Smith (1973) describes an interesting case of non-target-like chain-shift, whereby velarization of /t, d/ before laterals counterfeeds context-independent occlusivization of stridents to [t, d].

(5-34)

/pʌzəl/ → [pʌdəl] (* → [pʌgəl])
/pʌdəl/ → [pʌgəl]
/pɪkəl / → [pɪkəl]

As it is evident from the data in (5-34), underlying velars surface faithfully, and underlying /d/ is velarized to [g]. In its turn, /d/ derived by occlusivization does not undergo velarization process. In the light of rule-based approaches, in this case velarization does not apply to derived /d/, because its target is introduced too late in the derivation. Therefore, we are looking at the case of counterfeeding-on-focus. Recall from Chapter 3 that counterfeeding-on-focus is problematic for classic OT. OT-CC, on the other hand, claims to accommodate counterfeeding opacity. In the tableau below it is demonstrated how OT-CC deals with this case:

(5-35) *Spontaneous chain-shift in OT-CC* (adapted from Wolf 2008:355)

/pʌzəl/	***z**	**Ident(contin)**	**Prec(Ident(place), Ident(contin))**	***TL**	**Ident(place)**
< pʌzəl > *<Ø>*	*!			*	
→ < pʌzəl , pʌdəl > *<ID(contin)>*		*	*	*	
< pʌzəl , pʌdəl , pʌgəl > *<ID(contin), ID(place)>*		*	**!		*

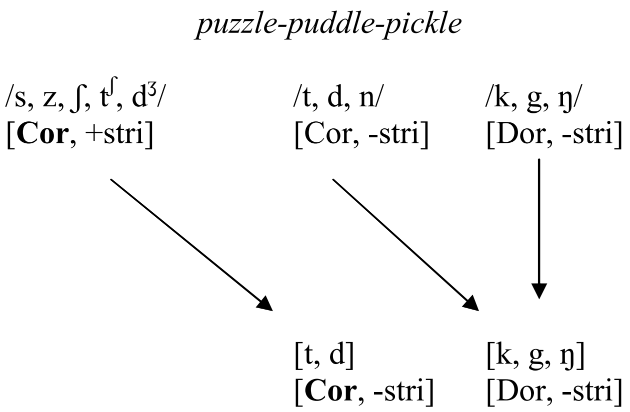
The metaconstraint on the ranking of Prec requires that Prec(Ident(place), Ident(contin)) should be ranked below Ident(contin). In order for Prec to favour the opaque candidate, it must also be ranked above its worst violation, i.e. above *TL. The above ranking yields the correct output, but it is also conceptually problematic.

McCarthy (2007:119) says that “if Prec constraints are universally present in grammars, then it is reasonable to assume that they are initially ranked at the bottom of the hierarchy, below even the faithfulness constraints”. Such ranking would also ensure that the metaconstraint is observed. If we assume that at the initial state Prec constraints occupy the lowest stratum in the hierarchy, then according to the provisions of BCD algorithm (see Prince & Tesar 2004 and Chapter 2 hereof), Prec (Ident(place), Ident(contin)) can only come to dominate *TL in response to the positive evidence. In our case, the learner clearly has no access to such positive evidence, since neither occlusivization of stridents nor velarization of stops are active in the target language (i.e. English).

Therefore, if we want this analysis to hold, the only way is to assume that the learners have the innate bias for ranking Prec constraints high. The idea of the innate ranking bias is nothing new: recall from Chapter 2 that innate domination of Markedness over Faithfulness is assumed in order to rule out excessively permissive grammars. Developing this idea, Wolf (2008:355) suggests that in order to accommodate child production data, Prec constraints should be ranked above markedness constraints at the initial state. By transitivity it would also mean that Prec constraints are ranked above faithfulness constraints, which contradicts the metaconstraint on ranking. In fact, as we will see below, it would suffice to say that Prec constraints are ranked above Faithfulness constraints at the initial state. Prec can come to dominate Markedness in the course of acquisition due to the markedness demotion in response to the positive evidence. In any case, ranking Prec constraints above Faithfulness violates the metaconstraint.

According to Wolf(2008:355), there exists an alternative strategy for dealing with developmental chain-shifts, i.e. positional faithfulness proposed by Jesney 2005 (recall also the discussion on zero-terminating chain-shifts from Chapter 4). According to Jesney(2005:59), the immunity of target /z/ to the general “d to g process” triggered by *TL is the consequence of “preferential feature preservation”, enforced by special Identity constraints. That is, input /z/ is said to contain a particularly well-formed feature combination [coronal, +strident]. Privileged relationship that holds between coronality and stridency has the effect that the feature [coronal] is preserved only when it is associated with [+strident] feature in the input.

(5-36) *Preferential preservation of [coronal] on input [+strident] segments* (from Jesney 2005:63)



(5-37) *Required ranking for the puzzle-puddle-pickle chain shift* (from Jesney (2005:87))

*[+strident], IdentCoronal/[+strident] >> *TL >>IdentCoronal/[-strident]

As noted by Wolf (2008:356), Jesney's approach to child chain shifts (combined with the assumption that Prec constraints are bottom-ranked in the initial state) predicts, *inter alia*, that chain shifts (counterfeeding-on-focus) are the only type of spontaneous opacity that should arise. Wolf continues to say that the discovery of cases of spontaneous counterbleeding would demonstrate the need for Prec constraints being high-ranked in the initial state.

The cases of spontaneous counterbleeding are, in fact, well attested. Barlow & Keare (2008:84), for example, describe the case of non-target-like opacity whereby the process of vowel-lengthening before voiced obstruents is counterbled by word-final voice neutralization.

(5-38) *Spontaneous counterbleeding* (adapted from Barlow & Keare 2008:84)

/dɔg/	→	[dɔːk]	'dog'	cf.	/dɔgi/	→	[dɔːgi]	'doggie'
/dʌk/	→	[dʌk]	'duck'	cf.	/dʌki/	→	[dʌki]	'duckie'

It is needless to say that preferential faithfulness cannot handle overapplication. OT-CC analysis of counterbleeding, in its turn, requires that relevant Prec constraint should dominate some faithfulness constraint. Since the learner lacks positive evidence necessary to establish such ranking, we have no choice but to assume that the ranking is due to the innate bias in conjunction with BCD.

The need for Prec constraints being high-ranked at the initial state has already been contemplated by scholars working on language acquisition. Thus, Dinnsen & Farris-Trimble (2008:115) come to the conclusion that “the fact that opacity effects emerge naturally in the course of early acquisition suggests that in the initial-state, Prec constraints are ranked relatively high among the faithfulness constraints”. The fact that Prec comes to dominate markedness constraint (in our case *TL) in the absence of any positive evidence, has been attributed to the “imperfect partial learning” (see Dinnsen & Farris-Trimble 2008 and references therein). In their analysis, Dinnsen & Farris-Trimble 2008 adhere to the original proposal by McCarthy 2007, whereby the ranking of Prec is restricted by the metaconstraint. However, above we have seen that there exist solid reasons in favor of the discarding of the metaconstraint on the ranking of Prec. Once the metaconstraint is discarded, nothing forces Prec to be low-ranked at the initial state. Below we will outline our proposal concerning the status of Prec constraints in more detail.

5.3 The status of Precedence constraints

If we assume that the innate bias for Markedness >> Faithfulness ranking is justified, discarding the metaconstraint leaves us with two options with the respect to the default ranking of Precedence constraints.

(5-39) *Hypothesized Initial State*

a. Precedence >> Markedness >> Faithfulness
b. Markedness >> Precedence >> Faithfulness

Let’s consider these options in turn. Prec (A,B) can be satisfied in three cases: first, by the faithful candidate that violates neither A nor B; second, by the candidate that violates only A; third, by the candidate that violates both A and B in the order prescribed by the relevant Prec constraint. We will illustrate the consequences of Prec being high-ranked with the example of counterfeeding opacity. If Prec is ranked above markedness at the initial state, the most faithful candidate will be chosen as optimal.

(5-40) *Spontaneous chain-shift in OT-CC with undominated Prec*

/pʌzəl/	Prec(ID(place), ID(contin))	*z	*TL	ID(place)	ID(contin)
☞< pʌzəl > *<Ø>*		*	*		
☹< pʌzəl , pʌdəl > *<ID(contin)>*	*		*		*
< pʌzəl , pʌdəl , pʌgəl > *<ID(contin), ID(place)>*	*!*			*	*

This is clearly not the result that we want, due to two reasons. First, it is inconsistent with the attested early production data. Second, and most important, if the output of child's grammar equals the adult output no learning can take place: BCD will converge on the incorrect and partial ranking Prec >> *z, *TL >> Id(place), Id(contin).

This ranking hypothesis becomes even more problematic if we assume, following McCarthy 2007, that Prec constraints are innate. It would mean that the uppermost stratum would contain not only Prec(A,B), but also Prec(B,A). Working together, the two Prec constraints would eliminate all candidates but a faithful parse that always vacuously satisfies Prec.

Option in 5-39 (b) is much more plausible. Prec can be viewed as some sort of a quasi-faithfulness constraint, because it is always vacuously satisfied by the faithful candidate. Ranked immediately below markedness, Prec constraints will jointly favor the most faithful of the unmarked candidates. In classic OT, this is exactly the job done by regular faithfulness constraints at the initial state. Therefore, Prec constraints ranked as shown in 5-39 (b) will not interfere with the evaluation process. At some point of the acquisition the learner is supposed to notice that the intended winner (adult form) violates some high-ranked markedness constraint. In the face of such positive evidence she will be prompted to demote the relevant markedness constraint. When Prec constraint comes to dominate markedness, opacity effects might arise (though it's not always that they do, as we will see in Chapter 6).

Another issue that is worth considering is whether Prec constraints are innate or constructed. In the original proposal, McCarthy (2007:119) in principle allows for both options, with a remark that the issue is not very important due to the narrow range of logically possible Prec constraints. However, if we allow Prec constraints to be relatively high-ranked at the initial state, we also endow them with the power to crucially affect the evaluation process. It is needless to say that Prec(A, B) alone and Prec (A, B) working in conjunction with Prec(B,A) will favor different candidates. Also, if we allow all logically possible Prec constraints to be present in the hierarchy, the evaluation process is at risk of being crucially affected by a large body of completely irrelevant constraints like Prec(C,B), Prec(D,B) etc.

Therefore, the assumption that Prec constraints are constructed by the learner in the process of acquisition seems justified. But it raises another important question: what prompts the learner to construct Prec constraints? In his brief discussion of learnability issues in OT-CC, McCarthy (2007:119) remarks that Prec constraints might be constructed when the learner faces “otherwise refractory data”, meaning that the Prec constraints are constructed as a ‘last resort’ when simple reranking of faithfulness and markedness cannot get the intended winner to be chosen as optimal. The main counterexample to this approach is the very existence of spontaneous opacity effects: the child who is trying to acquire transparent generalizations will never come across contradictory data, and therefore she will never have a reason to construct Prec constraints. So what is the trigger?

We propose that Prec constraints are constructed when the learner detects that the rLUMSeq of at least one of the candidates is not empty. It would be probably naïve to assume that of all logically possible (usually two) options, exactly the constraint that will ‘save the day’ is constructed. Therefore, we propose that once the learner detects that some candidate has a non-empty rLUMSeq, all logically possible constraints are constructed. The further course of events depends on whether the learner is trying to acquire opaque or transparent generalization. In the former case, the learner will place newly-constructed Prec constraints into the hierarchy one by one. The Prec constraint that will resolve the conflict in the data will be kept, while all the others will be discarded and will not participate in the further ranking. When the learner is trying to acquire transparent generalization, there is no contradiction in the data to start with, so any Prec constraint will ‘fit’. Which one is chosen, then, is purely a matter of chance. The choice, however accidental it might be, determines the intermediate stages the child will go through on her way to the adult output. In some cases, as we will see in Chapter 6, the child who opted for Prec_1 will go through the ‘opaque’ stage, while the outputs of the child who opted for Prec_2 will be consistently transparent. Thus we predict that spontaneous opacity effects are not obligatory, and that children having chosen different Prec constraint will go through different intermediate stages on their way to the adult output.

5.4. Summary

In this chapter we have considered three pieces of evidence that support discarding the metaconstraint on the ranking of Precedence constraints: obligatorily counterbleeding processes, derived environment effects and spontaneous opacity. These data prompted us to question the status of Prec constraints and introduce certain changes into the original proposal by McCarthy 2007. It goes without that saying that extensive research is necessary to see if they are empirically adequate. However, in the following chapters we will demonstrate that our assumptions about Precedence constraints make it possible to model both spontaneous opacity effects and the acquisition of target-like opacity in the light of OT-CC. Importantly, as it will be shown, the acquisition model that we propose is fully compatible with BCD algorithm.

Chapter 6: Spontaneous Opacity in Acquisition

6.1 Introduction

In the previous chapter we have reviewed typological evidence that prompted us to alter some basic assumptions about the status of Precedence constraints proposed by McCarthy 2007. Thus, we have demonstrated the necessity to discard the metaconstraint on the ranking of Prec. Computability considerations led us to assume that Prec constraints are constructed, rather than innate. In addition, the early production data have led us to propose that Prec constraints, when constructed, are placed in the middle stratum of the hierarchy. In this chapter we will demonstrate that our assumptions about Prec constraints make it possible to account for the emergence and subsequent disappearance of overapplication and underapplication opacity effects in developing grammars. In addition, we will see if OT-CC can provide any insight into such acquisition-related puzzles as U-shaped learning and cross-subject variation in production data.

6.2 Spontaneous Counterbleeding

In this section we will analyze emergent counterbleeding opacity effects demonstrated by the learners acquiring transparent generalizations. We will also address the problem of variation in early production data, and try to explain why some children demonstrate spontaneous opacity effects, while others have consistently transparent outputs.

6.2.1 Learner A: overapplication effects

Barlow & Keare 2008 present a case of spontaneous counterbleeding opacity, whereby children neutralize voicing contrast on syllable-final obstruents while still maintaining vowel lengthening before underlyingly voiced obstruents. Relevant child data are provided below.

(6-1) *Distribution of voiced and voiceless obstruents* (adapted from Barlow & Keare 2008:84)

a. Voiced and voiceless obstruents occur word-initially and intervocalically

[dɔ:k]	‘dog’	[t^{h}ʌ:p]	‘tub’
[bi:ts]	‘bridge’	[fis]	‘fish’
[wɛ:bi]	‘web (dim.)’	[soʊpi]	‘soap (dim.)’
[tʃi:zi]	‘cheese (dim.)’	[fisi]	‘fish (dim)’

b. Voice contrast is neutralized word-finally

[woʊ:p]	‘robe’	[soʊp]	‘soap’
[dɔ:k]	‘dog’	[dʌk]	‘duck’
[dʌ:f]	‘glove’	[wif]	‘leaf’
[tʃi:s]	‘cheese’	[ʤus]	‘juice’

(6-2) *Morphophonemic alternations for target morpheme-final voiced obstruents* (adapted from Barlow & Keare 2008:85)

[mʌ:t]	‘mud’	[mʌ:di]	‘mud (dim.)’
[dɔ:k]	‘dog’	[dɔ:gi]	‘dog (dim.)’
[dʌ:f]	‘glove’	[gʌ:vi]	‘glove (dim.)’
[tʃi:s]	‘cheese’	[tʃi:zi]	‘cheese (dim.)’

Before we analyze child production at this stage, we have to make certain assumptions about the state of the child’s grammar at this point. Errors in production indicate that the child is still struggling with language-specific phonotactics. Productive knowledge of voicing patterns also indicates that lexical representations of voiced and voiceless obstruents are adult-like (Barlow & Keare 2008:85). We also assume that at this point of acquisition the child has already departed from the identity map with respect to vowel length, having observed the fact that the length of a vowel is always predictable from the quality of the following consonant. In other words, the underlying representations the child has internalized are completely adult-like at this point in acquisition (i.e. /dɔg/ for [dɔ:g] etc.)

Provided that our above assumptions are correct, the error-pattern we are looking at is a case of counterbleeding opacity, whereby vowel-lengthening applies out of context, i.e. before a voiceless consonant. This counterbleeding opacity is spontaneous, or non-target-like, because neither lengthening of vowels before voiceless consonants nor final devoicing happen in the variety of English that the child is trying to acquire.

(6-3) *Spontaneous counterbleeding* (adapted from Barlow & Keare 2008)

a. /dɔg/ → [dɔːk] ‘dog’
b. /dɔgi/ → [dɔːgi] ‘dog(dim.)’
c. /dʌk/ → [dʌk] ‘duck’
d. /dʌki/ → [dʌki] ‘duck (dim.)’

In (6-3) we can see that the learner lengthens a vowel followed by the underlyingly voiced obstruent, and at the same time devoices the triggering obstruent. Below we will analyse this case of overapplication in the light of OT-CC.

Following Barlow & Keare (2008:87), we assume that the following constraints are relevant for the analysis:

(6-4)

***V̆C̬** – short vowels before voiced obstruents are prohibited.
***C̬#** – voiced coda obstruents are prohibited.
ID[length] – input and output segments should have identical specifications for [length]
ID[voice] – input and output segments should have identical specifications for [voice]

Barlow & Keare 2008 present the analysis of this case of spontaneous counterbleeding along the lines of OT-CC. Consistent with the original proposal by McCarthy (2007:119), they assume that precedence constraints are innate, and that at the initial state they are ranked at the bottom of the hierarchy. Furthermore, their analysis respects the metaconstraint on the ranking of precedence constraints. The analysis is given below, where Prec stands for Prec (ID[length], ID[voice]):

(6-5) *Spontaneous counterbleeding in OT-CC* (from Barlow & Keare 2008:88)

/dɔg/ ‘dog’	***V̆C̬**	***C̬#**	**ID[voice]**	**Prec**	**ID[length]**
a. <dɔg> faithful candidate	*!	*			
b. <dɔg, dɔːg> ID[length]		*!			*
c. ☞ <dɔg, dɔːg, dɔːk> ID[length], ID[voice]			*		*
d. <dɔg, dɔk> ID[voice]			*	*!	

In Tableau (6-5) above, the faithful candidate violates both high-ranked markedness constraints, and is therefore eliminated. Candidate (b), the intended winner, also violates a high-ranked markedness constraint against voiced codas. Transparent candidate (d) incurs a fatal violation of the Prec constraint, and thus opaque candidate (c) is correctly predicted to win.

Though the ranking illustrated above correctly accounts for the data, it raises a familiar question first sounded by Wolf (2008:355). What evidence do learners have for ranking Prec constraint that high? The answer is: none. On the basis of the positive evidence from the target-language, the learner must conclude that in order to get the optimal candidate (here: [dɔ:g]) to win, all is needed is the demotion of *C̬# below ID[voice]. The ranking shown in Tableau (6-5) could have resulted only if the learner maliciously intended to derive the 'illegal' opaque output. Acquisition algorithms available to date do not provide for such a possibility, and therefore the above analysis cannot be taken to reflect a plausible learning situation.

Below we present the acquisition model based on BCD (see Prince & Tesar 2004) and Minimal GLA (see Boersma 2008). In accordance with these works, we assume that at the initial state markedness constraints dominate faithfulness constraints. Based on our discussion in Chapter 5, we assume that a precedence constraint, when constructed and placed into the hierarchy, occupies a stratum immediately below markedness and above faithfulness. In addition to the markedness constraints suggested by Barlow & Keare 2008, we also postulate a context-free markedness constraint against long vowels, *LongV, ranked below the context-sensitive *V̆C̬. Therefore, the initial state is as shown below:

(6-6) *Ranking at the Initial State*

*V̆C̬ , *C̬# >>*LongV >> *Prec Constraint* >> ID[length], ID[voice]

Recall that in error-driven constraint demotion algorithms (see Tesar 1995, Prince & Tesar 2004, Boersma 2008), a suboptimal form or 'loser' is an output of the learner's current grammar. In OT-CC, however, suboptimal competing candidates are generated by the learner on the basis of the input and current constraint ranking. This solves the problem of the potentially infinite search space of competing output forms (Tesar 1995) and, combined with

Minimal GLA (Boersma 2008), ensures that the informative loser is always available for the language learner.

First the learner has to construct potential outputs for the input /dɔg/ on the basis of her current grammar.

(6-7) *Candidate chains for the input /dɔg/*

a. <dɔg>
b. <dɔg, dɔk>, because *C̬# >> ID[voice]
c. < dɔg, dɔ:g>, because *V̆C̬ >> ID[length]
d. <dɔg, dɔ:g, dɔ:k>, *V̆C̬ >> ID[length], and *C̬# >> ID[voice]

(6-8) *Candidates for the input /dɔg/ ordered as (in, out, L-set, rLUMseq)*

a. (/dɔg/, dɔg, Ø,Ø)
b. (/dɔg/, dɔk, ID[voice], Ø)
c. (/dɔg/, dɔ:g, ID[length], Ø)
d. (/dɔg/, dɔ:k, {ID[length], ID[voice]}, < ID[length], ID[voice]>)

The learner detects that the candidate set contains a candidate whose rLUMseq is not empty. This observation prompts the learner to construct Prec constraints: Prec(ID[length], ID[voice]) and Prec (ID[voice], ID[length]). As we mentioned in the previous chapter, what constraint gets chosen is purely incidental in this case. However, for the sake of illustration we will consider both scenarios in turn.

Suppose the learner has chosen Prec (ID[length], ID[voice]) and inserted it into its designated slot in the hierarchy. Then, in accordance with Minimal GLA (Boersma 2008), the learner has to choose some random total ranking of constraints consistent with her current stratified grammar and compute the output of her grammar. She realizes that it differs from the optimal output. This is illustrated in the Tableau (6-9) below.

(6-9) *Initial stage: transparent outputs*

/dɔg/	***V̆C̬**	***C̬#**	***LongV**	**Prec (ID[length], ID[voice])**	**ID[length]**	**ID[voice]**
a. dɔg	*	*!				
💣b. dɔk				*		*
☹c. dɔ:g		*!	*		*	
d. dɔ:k			*!		*	*

In Tableau (6-9) above, faithful candidate (a) is ruled out by high-ranked markedness constraint banning short vowels followed by voiced obstruents. The intended winner in (c) incurs a fatal violation of the markedness constraint against voiced codas, while opaque form (d) crucially violates *LongV. Thus, the transparent unmarked candidate in (b) is judged optimal.

Having computed the output of her current grammar, the child detects that it is different from the adult output: the child's grammar produces the output [dɔk], while predicting 'target' output [dɔ:g] to be suboptimal.

The learner constructs winner-loser pairs for the input /dɔg/, and performs the demotions necessary to make the form [dɔ:g] the winner. Constraint demotion is carried out in accordance with the BCD algorithm (see Prince & Tesar 2004).

(6-10)

/dɔg/	***V̆C̬**	***C̬#**	***LongV**	**Prec (ID[length], ID[voice])**	**ID[length]**	**ID[voice]**
a. dɔg	*	*				
c. dɔ:g		✪	✪		✪	
b. dɔk				*		*
c. dɔ:g		✪	✪		✪	
d. dɔ:k			*		*	*
c. dɔ:g		✪	✪		✪	

(6-11)

/dɔg/	***V̆C̬**	***C̬#**	***LongV**	**Prec (ID[length], ID[voice])**	**ID[length]**	**ID[voice]**
a. dɔg <<c. dɔ:g	W		L		L	
b. dɔk <<c. dɔ:g		L	L	W	L	W
d. dɔ:k <<c. dɔ:g		L				W

Since no ranking of markedness constraints can reconcile the learner with the data, the learner has to demote some markedness constraint. In accordance with Minimal GLA (Boersma 2008), the demotion is gradual, meaning that one constraint can be demoted by one stratum at a time. The first demotion of *LongV will result in the following hierarchy:

(6-12) *Intermediate stage: opaque outputs*

/dɔg/	***V̆C̬**	***C̬#**	**Prec (ID[length], ID[voice])**	***LongV**	**ID[length]**	**ID[voice]**
a. dɔg	*!	*				
b. dɔk			*!			*
☹c. dɔ:g		*!		*	*	
💣d. dɔ:k				*	*	*

In Tableau (6-12) above, candidate (a) is ruled out by the high-ranked markedness constraint banning short vowels before voiced obstruents. Candidate (b) is eliminated due to the violation of the Prec constraint. The intended winner, candidate (c) fatally violates the still-high-ranked *C̬#. Therefore, opaque candidate (d) is selected as an optimal output.

On the basis of winner-loser pairs, the learner again makes necessary adjustments to her grammar (gradually), until she arrives at the ranking where ID[voice] dominates *C̬#. At this point the learner has converged on the target grammar[8].

---

[8] Note that since *C̬# is dominated by ID[voice], forms *[dɔk] and *[dɔ:k] are no longer in the candidate set.

(6-13) *Final stage: transparent outputs*

/dɔg/	***V̌C̬**	**Prec (ID[length], ID[voice])**	***LongV**	**ID[voice]**	***C̬#**	**ID[length]**
a. dɔg	*!				*	
☞b dɔ:g			*		*	*

In Tableau (6-13) above, candidate (a) is ruled out by a high-ranked markedness constraint *V̌C̬. Therefore, candidate (b) is correctly selected as an optimal output, i.e. the child has learned the adult ranking.

It is probably worth highlighting that the final grammar in Tableau (6-13) is restrictive and also fully consistent with the Richness of the Base principle, i.e. it maps any input onto some language-legal output. Thus, the hypothetical input /dɔ:g/ is correctly mapped to [dɔ:g], while hypothetical /dɔ:k/ will be mapped to [dɔk].

It is also noteworthy that the input form /dɔgi/ is correctly mapped to [dɔ:gi] throughout the learning, which is consistent with the attested data.

(6-14) *Candidate chains for the input /dɔgi/:*

a. < dɔgi >
b. < dɔgi, dɔ:gi>, because *V̌C̬ >> ID[length]

(6-15) *Candidates for the input /dɔgi/ ordered as (in, out, L-set, rLUMseq)*:

a. (/dɔgi/, dɔgi, Ø,Ø)
b. (/dɔgi/, dɔ:gi, ID[length], Ø)

(6-16)

/dɔgi/	***V̌C̬**	**Prec (ID[length], ID[voice])**	***LongV**	***C̬#**	**ID[length]**	**ID[voice]**
a. dɔgi	*!					
☞b. dɔ:gi			*		*	

In this subsection we have seen that OT-CC correctly accounts for emergence and subsequent disappearance of overapplication effects in developing grammars. In the next subsection we will see if our theory is enough to account for the variation that is said to be a characteristic feature of early production data.

6.2.2 Learner B: alternative acquisition path

It has been observed that not all children go through the ‘opaque’ stage on their way to adult grammar: some learners (see Dinnsen et al. 2000) consistently produce transparent outputs. According to Dinnsen et al. (2000:337), “such variability is a hallmark of developing systems and is in no way atypical”. It goes without saying that a good acquisition model should be able to account for this state of affairs.

In the previous subsection we have demonstrated that OT-CC is able to account for the emergence and disappearance of overapplication opacity effects in developing grammars. In this section we will show that though OT-CC predicts (in full consistency with the attested data) that opacity effects might arise in the course of acquisition of transparent generalizations, it does not claim that such effects must and will arise.

Recall from Chapter 5 that Prec constraints are constructed by the learner when she detects that the rLUMSeq of some candidate is not empty. We have also assumed that the learner constructs all logically possible Prec constraints (usually ‘all’ means ‘two’). The further course of events depends on whether the learner is trying to acquire an opaque or a transparent generalization. In the former case, the learner will place newly-constructed Prec constraints into the hierarchy one by one. The Prec constraint that will resolve the conflict in the data will be kept, while all the others will be discarded and will not participate in the further ranking. When the learner is trying to acquire a transparent generalization, there is no contradiction in the data to start with, so any Prec constraint will ‘fit’. Which one is chosen is a matter of chance. The choice, however accidental it might be, determines the intermediate stages the child will go through on her way to the adult output. In some cases, the child who opted for Prec_1 will go through the ‘opaque’ stage, while the outputs of the child who opted for Prec_2 will be consistently transparent.

In Subsection 6.2.1 we have shown that the child who has demonstrated overapplication opacity effects while trying to acquire transparent generalization of English must have chosen Prec (ID[length], ID[voice]). In this section we will consider the alternative scenario and see what intermediate stages the learner would have gone through had she opted for Prec (ID[voice], ID[length]) instead.

The postulation of the ‘alternative’ Prec constraint, of course, will not influence the outcome of the learning process: the learner will still be able to eventually converge on the correct ‘adult’ output. However, the learning path of the hypothetical Learner B is predicted to be different from that of Learner A illustrated above. Note also that Learner B has exactly the same candidate set as Learner A, because Prec constraints do not participate in the construction of candidate chains.

Having placed Prec (ID[voice], ID[length]) into the hierarchy, Learner B computes the output of her current grammar. This is illustrated in Tableau (6-17) below.

(6-17)

/dɔg/	***V̆C̬**	***C̬#**	***LongV**	**Prec (ID[voice], ID[length])**	**ID[length]**	**ID[voice]**
a. dɔg	*!	*				
💣b. dɔk						*
☹c. dɔːg		*!	*	*	*	
d. dɔːk			*!	**	*	*

In Tableau (6-17), fully faithful candidate (a) violates both high-ranked markedness constraints. The intended winner, candidate (c), violates the markedness constraint against voiced codas. The opaque candidate (d) incurs a fatal violation of the *LongV constraint. Therefore, transparent candidate (b) is selected as optimal. Note that transparent candidate (b) harmonically bounds opaque candidate (d) – this ensures that Learner B will consistently have transparent outputs in the course of acquisition.

Having detected that the output of her grammar differs from the adult output, Learner B constructs winner-loser pairs for the input /dɔg/.

(6-18)

/dɔg/	***V̌C̬**	***C̬#**	***LongV**	**Prec (ID[voice], ID[length])**	**ID[length]**	**ID[voice]**
a. dɔg	*	*				
c. dɔ:g		✪	✪	✪	✪	
b. dɔk						*
c. dɔ:g		✪	✪	✪	✪	
d. dɔ:k			*	* *	*	*
c. dɔ:g		✪	✪	✪	✪	

(6-19)

/dɔg/	***V̌C̬**	***C̬#**	***LongV**	**Prec (ID[voice], ID[length])**	**ID[length]**	**ID[voice]**
a. dɔg <<c. dɔ:g	W		L	L	L	
b. dɔk <<c. dɔ:g		L	L	L	L	W
d. dɔ:k <<c. dɔ:g		L		W		W

Now the learner demotes the constraints violated by the intended winner in accordance with the BCD algorithm (Prince & Tesar 2004). In accordance with Minimal GLA (see Boersma 2008), the demotion should be minimal and gradual: one constraint can be demoted by one stratum at a time. At some point, Learner B will arrive at the target hierarchy as shown in Tableau (6-20) below[9].

(6-20)

/dɔg/	***V̌C̬**	**ID[voice]**	***C̬#**	***LongV**	**Prec (ID[voice], ID[length])**	**ID[length]**
a. dɔg	*!		*			
☞b. dɔ:g			*	*	*	*

---

[9] Note that since ID[voice] dominates ***C̬#,** forms *[dɔ:k] and *[dɔk] are no longer in the candidate set.

In Tableau (6-20), faithful candidate (a) violates a high-ranked constraint prohibiting short vowels before voiced obstruents. Therefore, candidate (b) is correctly chosen to be optimal: Learner B has successfully converged on the target grammar.

6.2.3 More cross-subject variation data

Dinnsen *et al.* 2000 provide data from two children suffering from phonological delay, which might be taken to exemplify both opaque and transparent acquisition paths. Thus, Child A showed evidence for the counterbleeding interaction of coda deletion and vowel lengthening before voiced consonants, whereby the vowel was lengthened even if the triggering consonant was deleted. Child C, in turn, demonstrated transparent interaction of vowel lengthening and coda deletion, whereby vowels preceding the deleted voiced consonant remained short. Consider the production data below:

(6-21) *Child A (age 7;2)* (from Dinnsen et al. 2000:325)

[kæ:]	‘cab’	[ka]	‘cop’
[kɪ:]	‘kid’	[pæ]	‘pat’
[dɔ:]	‘dog’	[dʌ]	‘duck’

(6-22) *Child C (age 3;10)* (from Dinnsen et al. 2000:338)[10]

[dɔ]	'dog'	[i]	'eat'
[dæ]	'dad'	[pel]	'plate'
[bɛ]	'bed'	[tʌ]	'truck'
[wɛ]	'red'		

If we assume that coda deletion is due to the markedness constraint NoCoda ranked above MAX, and vowel lengthening is due to *V̆C̬ ranked above Ident(length), the following candidate chains for the input /dɔg/ can be constructed:

---

[10] In fact, the situation is a bit more complex since Child C also deletes obstruents word-medially. Dinnsen et al. 2000 attribute it to the ranking NoCoda >>*VC >>MAX, where *VC is a markedness constraint against post-vocalic obstruents. In case of Child A, the ranking is NoCoda >> MAX >> *VC, to the effect that only coda obstruents are deleted. Since both children treat coda obstruents in exactly the same way, for the purposes of our analysis we will disregard the effects of *VC and assume the ranking NoCoda >> Max for both children.

(6-23) *Candidate chains for the input /dɔg/*

a. <dɔg>
b. <dɔg, dɔ>, because NoCoda >> MAX
c. < dɔg, dɔːg>, because *V̆C̬ >> ID[length]
d. <dɔg, dɔːg, dɔː>, *V̆C̬ >> ID[length], and NoCoda >> MAX

(6-24) *Candidates for the input /dɔg/ ordered as (in, out, L-set, rLUMseq)*

a. (/dɔg/, dɔg, Ø,Ø)
b. (/dɔg/, dɔ, MAX, Ø)
c. (/dɔg/, dɔːg, ID[length], Ø)
d. (/dɔg/, dɔː, {ID[length], ID[voice]}, < ID[length], MAX>)

In case of Child A, candidate (c) is selected as an optimal output at some point of acquisition. In case of Child C, it is candidate (b). Below we will show that our assumptions about the status of Prec constraints can account for the production data of both children.

Having constructed candidate chains and corresponding candidates, as shown in (6-23) and (6-24) above, both children will detect that rLUMSeq of one of the candidates is not empty. Therefore, the children will construct both logically possible Precedence constraints, in this case these are Prec(Ident(length), MAX)) and Prec(MAX, Ident(length)). At this point the acquisition path forks: Child A chooses Prec(Ident(length), MAX), while Child C chooses Prec(MAX, Ident(length)). As a result of the choice, Child A will go through the opaque stage, while Child C will have consistently transparent outputs.

(6-25) *Child A: opaque intermediate stage*

/dɔg/	***V̆C̬**	**NoCoda**	**Prec ID[length], MAX)**	***LongV**	**ID[length]**	**MAX**
a. dɔg	*!	*				
b. dɔ			*!			*
☹c. dɔːg		*!		*	*	
💣d. dɔː				*	*	*

In Tableau (6-25) above, candidate (a) is ruled out due to the violation of the high-ranked markedness constraint against short vowels followed by voiced obstruents. Unmarked candidate (b) fatally violates the Prec constraint, since MAX-violating LUM in its L-set is not preceded by ID[length]-violating LUM. The intended winner, candidate (c) fatally violates the high-ranked NoCoda. Therefore, overapplication candidate (d) is selected as optimal, in consistency with the attested data.

(6-26) *Child C: transparent intermediate stage*

/dɔg/	***V̆C̨**	**NoCoda**	**Prec (MAX, ID[length])**	***LongV**	**ID[length]**	**MAX**
a. dɔg	*!	*				
💣b. dɔ						*
☹c. dɔ:g		*!	*	*	*	
d. dɔ:			**!	*	*	*

The situation with Child C shown in Tableau (6-26) is somewhat different. Just like in case with Child A, candidates (a) and (c) are ruled out by the high-ranked markedness constraints. Opaque candidate (d) fatally violates the Prec constraint: ID[length]-violating LUM in its L-set is followed by MAX-violating LUM, which is the opposite of what the Prec constraint requires. Therefore, transparent candidate (b) is selected as optimal. Note also that opaque candidate (d) is harmonically bounded by (b), which means that Child C is predicted not to have opaque outputs at any point of acquisition.

In a nutshell, from our above analysis it follows that differences in production of Child A and Child C are due to the fact that these children follow *different acquisition paths*. However, this is not the only way to look at the data in question.

Dinnsen *et al.* 2000 analyze the production data of Child A and Child C in the light of Sympathy Theory (see McCarthy 1999; also see Chapter 3 hereof for discussion). Under their approach, differences in production between Child A and Child C are due to the fact that at the moment when the data were attested the children were at the *different stages of development*. This is schematically represented below:

(6-27) *Stages of development* (from Dinnsen et al. 2000:343)

Stage 1: Markedness >> Sympathy >> IO Faith
(e.g., Child C, transparent outputs)
Stage 2: Sympathy >> Markedness >> IO Faith
(e.g., Child A, opaque outputs)
Stage 3: Sympathy >> IO Faith >> Markedness
(e.g., archetypical fully developed language, transparent outputs)

According to Dinnsen *et al.* 2000, a sympathy constraint can affect evaluation only if it dominates some markedness constraint. Therefore, at the first stage of acquisition, the effect of the sympathy constraint is not visible, and the learner produces transparent outputs (according to Dinnsen *et al.* 2000, this stage is exemplified by Child C). Then, on the basis of positive evidence, markedness constraints are gradually demoted, so that at some point they are dominated by the sympathy constraint. This is when the opaque outputs are produced (this is allegedly the stage Child A is at). Finally, markedness constraints get demoted below IO faithfulness constraints, rendering sympathy constraints inert.

The insight expressed by Dinnsen *et al.* 2000 is largely consistent with the OT-CC-based acquisition model outlined in this section. Thus, above we have seen that Prec, just like sympathy, can influence the outcome of evaluation if it is ranked above some markedness constraint. Just like sympathy, Prec can only come to dominate markedness if markedness is demoted on the basis of positive evidence. Just like sympathy, Prec constraints become inert if relevant markedness constraints get demoted below faithfulness constraints. Recall from Subsection 6.2.1 that at the initial stage of acquisition our Learner A had transparent outputs, and it was only at the later stages that opacity effects emerged. Therefore, just like Dinnsen *et al.* 2000, we could have claimed that differences in production between Child A and Child C are merely due to the fact that the children are at different stages of development.

However, certain considerations prompt us to be cautious with drawing conclusions. Firstly, the account presented in Dinnsen *et al.* 2000 predicts that opacity effects will obligatorily arise at a certain point in acquisition. Moreover, the 'opaque' stage must follow and not precede the 'transparent' stage. Intuitively, this claim seems to be too strong; however, scrupulous longitudinal studies are necessary to see if it is warranted. Secondly, though Child C (aged 3;10) is indeed younger than Child A (aged 7;2), the fact that both children suffer from phonological delay might prevent us from taking this particular case to exemplify continuity in phonological development. To this end, it would be especially interesting to carry out a dynamic investigation of several typically developing children and see if any correlation can be drawn between their development and emergence of opacity effects (if any). Thirdly, in Chapter 3 we have reviewed some reasons why Sympathy Theory is not a particularly successful theory of (spontaneous) opacity. Therefore, in the light of the abovementioned, our hypothesis is that differences in production between Child A and Child C are due to the fact that the children have postulated different Prec constraints in the course of the acquisition. The hypothesized learning progression for Child A and Child C is summarized below:

(6-28) *Hypothesized learning progression for Child A and Child C*

Child A

Early stage: transparent outputs
*V̆Ç̬ , NoCoda >> *LongV >> Prec ID[length],MAX) >> ID[length], MAX

Intermediate stage: opaque outputs
*V̆Ç̬ , NoCoda >> Prec ID[length],MAX) >> *LongV >> ID[length], MAX

Final stage: target-appropriate outputs
*V̆Ç̬ >> Prec ID[length],MAX) >> *LongV >> ID[length], MAX >> NoCoda

Child C

Early stage: transparent outputs
*V̆Ç̬, NoCoda >> *LongV >> Prec (MAX, ID[length]) >> ID[length], MAX

Intermediate stage: transparent outputs
*V̆Ç̬, NoCoda >> Prec (MAX, ID[length]) >> *LongV >> ID[length], MAX

Final stage: target-appropriate outputs
*V̆Ç̬ >>Prec (MAX, ID[length]) >> *LongV >> ID[length], MAX >> NoCoda

In this section we have demonstrated that OT-CC can successfully account for the emergence and subsequent disappearance of overapplication effects in developing grammars. Importantly, we have seen that though OT-CC predicts (in full consistency with the attested data) that opacity effects might arise in the course of acquisition of transparent generalizations, it does not claim that such effects must and will arise: whether or not the learner will have opaque outputs depends entirely on which Prec constraint is chosen. Thus, OT-CC predicts a certain degree of variation in early production data. In the next section we will see if the assumptions we have made and conclusions we have reached so far can help us to analyze emergent underapplication effects in developing grammars.

6.3 Spontaneous counterfeeding

Now we will apply the modified version of OT-CC to another example of spontaneous opacity attested by Smith (1973), which has already been briefly discussed above. Smith (1973) reports that his son Amahl demonstrated an interesting case of non-target-like counterfeeding, whereby velarization of /t, d/ before laterals counterfed context-independent occlusivization of stridents to [t, d]. In other words, the case in question is a chain-shift, whereby derived instances of /d/ are protected from velarization.

(6-29)

pʌzəl → pʌdəl *→ pʌgəl
pʌdəl → pʌgəl
pɪkəl → pɪkəl

We have already demonstrated in Chapter 3 that chain-shifts *per se* are problematic for classic OT. The analysis of developmental chain-shifts is even more challenging: one has to account not only for the opaque generalization as such, but also for its emergence and subsequent loss. Several OT-compatible analyses of developmental chain-shifts have been offered to date (see, *inter alia*, Wolf 2008, Jesney 2005). Below we will briefly consider an LC-based account proposed by Dinnsen *et al.* 2001, since it follows loosely the same logic as the OT-CC-based analysis we will develop in this section.

Although the formal tools used for the two analyses are very different (LC on the one hand, and OT-CC on the other), the key ideas behind the analyses are quite similar. Firstly, both Dinnsen *et al.* 2001 and the author of this thesis proceed from the assumption that underlying representations internalized by Amahl are target-appropriate (see Chapter 2 and references therein for the motivation behind this view). It follows, then, that errors in production, including opacity effects, can only result from constraint ranking. Secondly, just like Dinnsen *et al.* 2001, we assume that emergence and loss of error patterns should occur naturally, as a result of markedness demotion prompted by positive evidence. Thirdly, no additional assumptions are made about the nature of underlying representations or markedness and faithfulness constraints. Having said that, we can now consider the analysis presented in Dinnsen *et al.* 2001 in somewhat more detail.

In their analysis of Amahl's chain-shift, Dinnsen *et al.* 2001 appeal to the local conjunction of ID[manner]&ID[place] in order to rule out */pʌzəl/→[pʌgəl] mapping. In their account, (see Dinnsen *et al.* 2001:518), the proposed ranking for the early acquisition stage is the following:

(6-30)

Early stage: LC, *Fric, *dl >> *gl >> ID[manner], ID[place][11]

LC constraint can only affect the outcome of the evaluation if it is ranked at the topmost stratum. Provided that in this case LC is a conjunction of two faithfulness constraints, i.e. a faithfulness constraint itself, it is not quite clear how it ended up being that high ranked. Although Dinnsen *et al.* (2001:512) state that "the ranking of the complex constraint over the elementary constraints is presumed to be universal, following from the Elsewhere Condition and the special/general relation that holds among these constraints", the durable bias for ranking markedness over faithfulness (Prince & Tesar 2004) in conjunction with the special/general relation bias would result in the ranking where *Fric, *dl, *gl >> LC >> ID[manner], ID[place]. LC, then, can come to dominate markedness constraints only as a result of the constraint demotion. In addition, in Chapter 3 we have already addressed the question why local conjunction might have undesirable consequences for acquisition.

The chain-shift, however, is only a part of the famous 'puzzle-puddle-pickle' problem. Smith (1973) reports that at the later stage of development, Amahl for some time departs from his correct pronunciation of velar-lateral clusters and pronounces /k, g/ as [t,d] when they are followed by laterals. Consider the data below:

(6-31)

pʌzəl →pʌzəl
pʌdəl → pʌdəl
pɪkəl → pɪtəl

---

[11] According to Dinnsen et al. (2001:516), the ranking of *dl over *gl might be a result of a 'default preference for assimilation over dissimilation'.

This is an instance of so-called U-shaped learning, or, as defined in Stemberger, Bernhardt & Johnson (1999:1), “a developmental change in which there is a decrease in accuracy relative to the adult target”.[12] On the standard view, the fact that at the early stage of acquisition Amahl produced *pickle*-words target-appropriately is suggestive of the ranking where ID[place] >> *gl. The subsequent departure from correct pronunciation, therefore, is not only entirely unexpected, but also cannot be modelled by means of BCD algorithm, which does not allow for either faithfulness demotion or markedness promotion.

According to Dinnsen *et al.* (2001:516), the key to the solution of this problem is “the recognition that some target-appropriate productions can arise and be judged optimal even though the ranking does not conform to the target ranking” (Dinnsen *et al.* 2001:516). In other words, target-appropriate realization of *pickle*-words at the early stage is attributed to the ranking *dl >> *gl >> ID[place], while the /pɪkəl/ → [pɪtəl] mapping at the later stage is a result of the demotion of *dl below *gl.

According to Dinnsen *et al.* (2001:523), overgeneralization patterns in acquisition are predicted to arise “if and only if an error pattern is governed by two opposing and overlapping universal markedness constraints which are not dominated by an antagonistic faithfulness constraint”. Below we will demonstrate that our OT-CC-based analysis is fully consistent with this important insight.

For our analysis we will need the following constraints (following Dinnsen *et al.* 2001):

(6-32)

***TL** – avoid coronals before liquid consonants[13]
***KL** – avoid velars before liquid consonants
***z** – avoid fricatives
Ident(manner) – underlying segments and their output correspondents must have identical specifications for manner
Ident(place) – underlying segments and their output correspondents must have identical specifications for place

Following the large body of works (see, *inter alia*, Prince & Tesar 2004, Hayes 2004; also see Chapter 2 hereof for the discussion), we assume that markedness constraints dominate

---

[12] See Dinnsen et al. (2001) for the discussion on Duke-of-York Gambit in acquisition.

[13] Following Dinnsen et al. (2001) we assume that *TL constraint prohibits all coronals before liquids. However, the reformulation of this constraint to affect coronal stops only will not have any crucial consequences for our analysis.

faithfulness at the initial state. Prec constraints, when constructed, are placed into the middle stratum of the hierarchy. The initial stratified hierarchy, therefore, is as shown below:

(6-33)

*z , *TL, *KL >> *Prec* >> IdManner, IdPlace

Interestingly, under the assumption that all markedness constraints are mutually unranked and the violations incurred by the candidate are summed up ('pooling ties', see Tesar 1995), construction of harmonically improving candidate chains for the input /pʌdəl/, for example, on the basis of the ranking given in (6-33) is impossible[14]. Thus **<pʌdəl, pʌgəl > is not a valid chain, because the form /pʌgəl/, though it adds up a violation of Ident(place), is not harmonically improving relative to /pʌdəl/– it violates the markedness constraint *KL that is ranked as high as *TL.

Such situations are known to arise also in fully developed grammars, when conventional ranking argumentation cannot establish the mutual ranking of some markedness constraints, e.g. in cases when both constraints are satisfied by the winning candidate and therefore are not in conflict. In certain cases, the ranking of such markedness constraints is crucially necessary to construct candidate chains. McCarthy (2007:81) suggests that in such cases chain validity itself can serve as a ranking argument: if there is no evidence for mutual ranking of (A, B), and harmonic improvement requirement on chains requires that the candidate satisfying A is more harmonic than the candidate satisfying B, then A>>B.

Clearly, when applied to the problem of language acquisition this solution faces the risk of being circular: candidate chains depend on the current ranking, which, in turn, depends on the validity of candidate chains. Unlike in adult grammars, children have no particular reason to assume that < pʌdəl, pʌgəl > is to be a harmonically improving chain, since the mapping /pʌdəl/→[pʌgəl] is not the target of acquisition. The fact that this chain is harmonically improving should, therefore, be a mere consequence of the learner's grammar at some point of the acquisition. We can actually model this situation with the help of the tool that we have already discussed above.

---

[14] Under the assumption that *TL only affects coronal stops, the chain <pʌzəl , pʌdəl> would be impossible as well.

Recall from Chapter 2 our discussion of the Minimal GLA (Boersma 2008), according to which at every point of development the learner's grammar represents a total refinement consistent with the learner's current stratified grammar. In our case, the stratified grammar is such that all markedness constraints dominate all faithfulness constraints. The production data, in its turn, suggest that at least at some point in acquisition the total refinement is such that *z>>*TL>>*KL. Let us see how the acquisition progresses.

First, the learner gets the input[15] and constructs possible output candidates for this input on the basis of her current constraint ranking.

(6-34)

Input	Candidate Chains	Candidates
/ pʌdəl /		
	a. < pʌdəl >	a. (/pʌdəl /, pʌdəl, Ø,Ø)
	b. < pʌdəl , pʌgəl >	b. (/pʌdəl /, pʌgəl, {IdPlace}, Ø)
/pikəl/		
	a. < pɪkəl > [16]	a. (/pɪkəl /, pɪkəl, Ø,Ø)
/pʌzəl/		
	a. < pʌzəl >	a. (/pʌzəl /, pʌzəl, Ø,Ø)
	b. < pʌzəl , pʌdəl >	b. (/pʌzəl /, pʌdəl,{ IdManner}, Ø)
	c. < pʌzəl , pʌdəl , pʌgəl >	c. (/pʌzəl /, pʌgəl, { IdManner, IdPlace}, <IdManner, IdPlace>)

The learner detects that the rLUMSeq of one of the output candidates is not empty. Therefore, the learner postulates Precedence constraints: Prec(Ident(place), Ident(manner)) and Prec(Ident(manner), Ident(place)). As we have already mentioned above, in case of spontaneous opacity the choice of a Prec constraint is purely incidental. In our case, acquisition data suggest that Prec(Ident(place), Ident(manner)) was chosen.

(6-35)

Prec(Ident(place), Ident(manner)) – the violation of Ident(manner) must be preceded and must not be followed by the violation of Ident(place).

---

[15] Here we also assume that the child's input equals adult output, but since in this case adult output equals adult input the distinction is not visible.

[16] Note that the chain **< pɪkəl, pɪdəl> is invalid because *TL > *KL.

Having placed the Prec constraint into its designated slot in the hierarchy, the learner computes the output of her current grammar.

(6-36) *Initial state: transparent outputs*

1.

pʌzəl	*z	*TL	*KL	Prec(IdPlace, IdManner)	IdManner	IdPlace
☹a. pʌzəl	*!	*				
b. pʌdəl		*		*	*	
💣c. pʌgəl			*	**	*	*

2.

pʌdəl	*z	*TL	*KL	Prec(IdPlace, IdManner)	IdManner	IdPlace
☹a. pʌdəl		*!				
💣b. pʌgəl			*			*

3.

pikəl	*z	*TL	*KL	Prec(IdPlace, IdManner)	IdManner	IdPlace
a. pikəl			*			

In Tableau (6-36-1) above, candidate (a) is ruled out due to violations of high-ranked *z and *TL. Candidate (b) fatally violates *TL, and therefore candidate (c) is judged optimal. In Tableau (6-36-2), the intended winner in (a) loses to candidate (b) on high-ranked *TL. In Tableau (6-36-3), the faithful target-appropriate form [pikəl] is the only output candidate, and so it is predicted to surface. Thus, at the initial stage of acquisition, the learner's outputs are predicted to be transparent: both /z/ and /d/ surface as [g]. Smith (1973) does not report of Amahl's production data that corresponds to this stage. This might be due to several reasons: this stage might be pre-verbal, and therefore no production data could be attested. It might be also very brief, which made it difficult to 'catch' the right production. Also, the effects of this ranking might have been obscured by some other high-ranked markedness constraints, e.g. constraints against word-medial consonants or constraints against disyllabic words.

Having detected that his outputs for /pʌzəl/ and /pʌdəl/ differ from the adult outputs, the learner constructs winner-loser pairs. Consider Tableau (6-37) below:

(6-37)

	*z	*TL	*KL	Prec(IdPlace, IdManner)	IdManner	IdPlace
1./pʌzəl/						
b. pʌdəl < a. pʌzəl	L			W	W	
c. pʌgəl <a. pʌzəl	L	L	W	WW	W	W
2./pʌdəl/						
b. pʌgəl < a. pʌdəl		L	W			W

Where the output of the learner's current grammar is not identical with the optimal (adult) output, the learner demotes the highest-ranked constraint violated by the optimal output by one stratum. Demotion is minimal and gradual, in accordance with Minimal GLA (Boersma 2008). Since both *z and *TL prefer losers, the child demotes *z and *TL for one stratum, i.e. immediately below *KL. The mutual ranking of markedness is now the following: *KL >> *z >> *TL.

Once again, the learner has to construct candidate chains on the basis of this new ranking. The new set of candidate chains and their corresponding candidates is given in the table below:

(6-38)

Input	Candidate Chains	Candidates
/ pʌdəl /		
	a. <pʌdəl> [17]	a. (/pʌdəl /, pʌdəl, Ø,Ø)
/pikəl/		
	a. <pɪkəl>	a. (/pɪkəl /, pɪkəl, Ø,Ø)
	b. <pɪkəl, pɪtəl>	b. (/pɪkəl/, pɪtəl,{ Ident(place)}, Ø)
/pʌzəl/		
	a. < pʌzəl >	a. (/pʌzəl /, pʌzəl, Ø,Ø)
	b. < pʌzəl , pʌdəl >	b. (/pʌzəl /, pʌdəl,{ IdManner}, Ø)

Having constructed valid harmonically improving candidate chains, the learner repeats the evaluation.

---

[17] Note that chain <pʌdəl , pʌgəl> is invalid, because at this point *KL >> *TL

(6-39) *Early stage: transparent outputs*

1.

pʌzəl	*KL	*z	*TL	Prec(IdPlace, IdManner)	IdManner	IdPlace
☹a. pʌzəl		*!	*			
💣b. pʌdəl			*	*	*	

2.

pʌdəl	*KL	*z	*TL	Prec(IdPlace, IdManner)	IdManner	IdPlace
a. pʌdəl			*			

3.

pikəl	*KL	*z	*TL	Prec(IdPlace, IdManner)	IdManner	IdPlace
☹a. pikəl	*!					
💣b. pitəl			*			*

In Tableau (6-39-1) above, the intended winner in (a) loses to candidate (b) on high-ranked *z constraint. In Tableau (6-39-2), the target-appropriate form [pʌdəl] is predicted to surface since it is the only candidate. In Tableau (6-39-3), the intended winner in (a) is ruled by the high-ranked *KL. Once again the learner's outputs are transparent, that is, both /z/ and /g/ surface as [d]. The fact that this stage is unattested might be due to the reasons already enumerated above.

Since the learner's outputs still differ from adult output forms, constraint demotion continues. At some point *TL and *KL are demoted below the Precedence constraint. One of the random rankings consistent with the stratified hierarchy is as follows:

(6-40)

*z>> Prec(IdPlace,IdManner) >>*TL >>*KL >>IdManner, IdPlace

The learner again constructs output candidates on the basis of her new ranking and repeats the evaluation, as shown in Tableau (6-41) below:

(6-41) *Intermediate stage 1: z→d→g chain-shift*

	*z	Prec(IdPlace, IdManner)	*TL	*KL	IdManner	IdPlace
1./pʌzəl/						
☹a. pʌzəl	*!		*			
💣b. pʌdəl		*	*		*	
c. pʌgəl		**!		*	*	*
2./pʌdəl/						
☹a. pʌdəl			*!			
💣b. pʌgəl				*		*
3./pikəl/						
a. pikəl				*		

In Tableau (6-41), faithful candidate (1a) fatally violates the high-ranked markedness constraint *z. Candidate (1c) fares worse than (1b) on the Prec constraint, since Ident(manner)-violating LUM in its L-set is followed by Ident(place)-violating LUM. Candidates (2a) and (2b) both vacuously satisfy the Prec constraint, since neither of them violates Ident(manner). The intended winner in (2a), however, loses to candidate (2b) on the markedness constraint *TL. The target-appropriate candidate (3) is the only candidate possible under the current ranking, and therefore it is predicted to surface. In other words, Tableau (6-41) illustrates the solution of a famous puzzle-puddle-pickle problem, i.e. the chain-shift, whereby /z/→[d], /d/ → [g] and /g/ → [g].

Since the learner's outputs are still not target-appropriate, more demotions are necessary. Thus, *z gets demoted below Prec, while *TL gets demoted below *KL. Then the learner randomly chooses a total ranking consistent with his new stratified hierarchy. Suppose the ranking is as shown below:

(6-42)

Prec (IdPlace, IdManner) >>*z >> *KL >> *TL >> Ident (place) , Ident(manner)

Now the learner again constructs output candidates on the basis of her current grammar. The candidates will be exactly as those in Table (6-38) above. Having constructed the candidate set, the learner again computes the output of his grammar, and realizes that further demotions are needed.

(6-43) *Intermediate stage 2: U-shaped learning effects*

	Prec(IdPlace, IdManner)	*z	*KL	*TL	IdPlace	IdManner
1./pʌzəl/						
☞a. pʌzəl		*		*		
b. pʌdəl	*!			*		*
2./pʌdəl/						
☞a. pʌdəl				*		
3.pikəl						
💣a. pitəl				*	*	
☹b. pikəl			*!			

Tableau (6-43) above characterizes the stage where opacity effects are lost: the intended winner in (1a) is correctly predicted to surface, regardless the fact that *z still dominates both faithfulness constraints. Target appropriate form in (2) is judged optimal since it is the only candidate possible under the current ranking. The intended winner in (3b) loses to candidate (a) on undominated markedness constraint *KL. Thus, Tableau (6-43) illustrates the solution of the U-shaped learning problem, whereby the learner departs from once-correct production of *pickle*-words.

Having detected that the outputs of his grammar are different from the adult outputs, the learner will be motivated to demote the markedness constraints. At some point, positive evidence will lead the learner to demote both *TL and *KL below Ident(place). The resulting stratified grammar is as follows:

(6-44)

Prec (IdPlace, IdManner) >> *z >> IdPlace>> *TL , *KL >> IdManner

(6-45) *Final stage: target-appropriate outputs*

	Prec(IdPlace, IdManner)	*z	IdPlace	*TL	*KL	IdManner
1./pʌzəl/						
☞a. pʌzəl		*		*		
b. pʌdəl	*!			*		*
c. pʌgəl	*!*		*		*	*
2./pʌdəl/						
☞a. pʌdəl				*		
3.pikəl						
☞a. pikəl					*	

In Tableau (6-45) above, candidates (b) and (c) are ruled out by now-undominated Prec constraint. Therefore, the intended winner in (1a) is correctly selected as optimal even though *z is not dominated by faithfulness constraints. Target-appropriate forms in (2) and (3) are predicted to surface since no other candidates are possible under the current ranking. At this point the learner has converged, since his outputs are identical to adult outputs. The stratified hierarchy in (6-45) is totally refinable, i.e. any total ranking consistent with it will produce correct outputs (i.e. the mutual ranking of *KL and *TL at this point is not crucial). We might also assume that the learner is biased for demoting the constraints violated by the optimal candidate as low as possible, so the grammar might be further refined to the following:

(6-46) *Hypothesized adult grammar*

Prec (IdPlace, IdManner) >> IdPlace, IdManner >> *TL, *KL, *z

Note that in the grammar in (6-46), Prec constraint is inert: since all markedness constraints are dominated by the relevant faithfulness constraints, only faithful parses can be included into the candidate set.

Thus we have demonstrated that OT-CC can account for the emergence and subsequent loss of error patterns attested by Smith (1973). Our findings are summarized below:

(6-47) *Hypothesized learning progression for Amahl*

1. Initial state: transparent outputs *[unattested]*
*z>>*TL>>*KL >>Prec(ID(place), ID(manner)) >> ID(manner), ID(place)

2. Early stage: transparent outputs *[unattested]*
*KL >> *z >> *TL >> Prec(ID(place), ID(manner)) >> ID(manner), ID(place)

3. Intermediate stage 1: z→d→g chain-shift *[attested]*
*z >> Prec(ID(place), ID(manner)) >> *TL >> *KL >> ID(manner), ID(place)

4. Intermediate stage 2: U-shaped learning effects *[attested]*
Prec(ID(place), ID(manner)) >> *z >> *KL >>*TL >> ID(place), ID(manner)

5. Final stage: target-appropriate outputs *[attested]*
Prec(ID(place), ID(manner)) >> *z >> ID(place), ID(manner) >> *KL, *TL

In this section we have shown that it is possible to account for developmental chain-shifts in the light of OT-CC, while at the same time closely adhering to the widely held and well-substantiated assumptions about the nature of underlying representations, initial state and acquisition progression (see Chapter 2 and references therein). We have demonstrated that constraint rankings corresponding to the attested error patterns arise naturally in the course of acquisition, and the order in which they arise is fully consistent with the attested data.

6.4 Summary

In this chapter we have presented the OT-CC-based account of the emergence and subsequent loss of spontaneous counterbleeding and counterfeeding effects in developing grammars. Importantly, we have seen that though OT-CC predicts (in full consistency with the attested data) that opacity effects might arise in the course of acquisition of transparent generalizations, it does not claim that such effects must and will arise. Thus, our analysis predicts a certain degree of variation in early production data. The analysis we have provided is fully compatible with BCD algorithm (see Prince & Tesar 2004) and Minimal GLA (Boersma 2008). Besides, it follows the widely held and well-substantiated assumptions about the nature of underlying representations, initial state and learning progression. We have demonstrated that constraint rankings corresponding to the attested error patterns arise naturally in the course of acquisition, and the order in which they arise is fully consistent with the attested data. Besides, we have shown that there is a high degree of continuity in transition from one developmental stage to the next. Another obvious advantage of our analysis is that we have avoided the dichotomy problem: the same basic principles and the same formal tools are used to account for both overapplication and underapplication effects in developing grammars.

Although further evidence from early production is necessary to substantiate our claims pertaining to the learning progression, OT-CC-based account outlined in this chapter solves many problems inherent in the previous approaches to spontaneous opacity, while at the same preserving their merits.

Chapter 7: Acquisition of Target-Like Opacity

7.1 Introduction

In Chapter 3 we have already seen that opacity effects pose considerable challenges for OT in its classic ‘parallel’ formulation. In Chapter 4 we have outlined Optimality Theory with Candidate Chains, the novel approach to opacity proposed by McCarthy 2007. We have also demonstrated above that some changes to the status of Precedence constraints make it possible for OT-CC to handle non-target-like opacity effects that emerge in developing grammars. In this chapter we will see if OT-CC is capable of accounting for the acquisition of target-like opacity effects. To illustrate the learning progression, we have chosen the notorious case of ‘Canadian Raising’. As shown in (7-1), in CE intervocalic /t/ and /d/ are realized as a flap, which produces alternations like [fæt] cf. [færər]. Flapping also applies across the word-boundary, resulting in the mappings like /hi hɪt æn/ → [hi hɪɾ æn]. Also, in CE low diphthongs [aɪ] and [ɑʊ] surface before voiced segments, while high diphthongs [əi] and [ʌʊ] surface before voiceless segments. Consider the data below:

(7-1) *Alternations in Canadian English* (from Bermúdez-Otero 2003)

1. Flapping in CE: lax intervocalic /t/ and /d/ are realized as [ɾ]

a. [fæɾər]	*fatter*	cf. [fæt]	*fat*
b. [mæɾər]	*madder*	cf. [mæd]	*mad*
c. [hi hɪɾ æn]	*he hit Ann*	cf. [hɪt]	*hit*
d. [hi hɪɾ æn]	*he hid Ann*	cf. [hɪd]	*hid*

2. Diphthong Raising in CE: /aɪ/ and /ɑʊ/ are raised to [əi] and [ʌʊ] before a voiceless obstruent in the same foot

a. [nəɪf]	*knife*	cf. [naɪvz]	*knives*
[hʌʊs]	*house*	cf. [hɑʊzɪz]	*houses*
b. [‘səifn]	*syphon*	cf. [saɪ’fɑnɪk]	*syphonic*
[səit]	*cite*	cf. [saɪ’teɪʃn]	*citation*

Interaction of Flapping and Raising results in overapplication, e.g. in [ɹ**əiɾ**ɪŋ], where a raised diphthong surfaces before a flap. The rule-based analysis of CE counterbleeding is given below in (7-2):

(7-2) *Counterbleeding interaction of Flapping and Raising* (from Bermúdez-Otero 2003)

	writing	*riding*	*mitre*	*spider*
UR	/ɹaɪt-ɪŋ/	/ɹaɪd-ɪŋ/	/maɪtəɹ/	/spaɪdəɹ/
Raising	ɹəitɪŋ	---	məitər	---
Flapping	**ɹəiɾ**ɪŋ	ɹaɪɾɪŋ	m**əiɾ**əɹ	spaɪɾəɹ

Interpretation of CE data caused rather heated controversy among scholars. Some scholars (Idsardi 2005, Bermúdez-Otero 2003) recognized the data in (7-2) as a true case of phonological opacity, where raising of underlying /aɪ/ to [əɪ] seems to have overapplied before a flap, e.g. in [ɹəiɾɪŋ]. Yet some others (Kaye 2009, Fruehwald 2007, Hayes 2004, Mielke *et al.* 2003, Vance 1987) suggested that the distinction between [aɪ] and [əi] might be phonemic (at least for some speakers), i.e. diphthongs in *rider* and *writer* are not derived from the same underlying source.

For the purposes of this thesis, we interpret CE data in (7-2) above as an example of counterbleeding opacity, following Bermúdez-Otero 2003. Should the opposite be ever proved beyond reasonable doubt, it would not diminish theoretical and empirical merits of OT-CC: the analysis below shows that acquisition of opacity is not a principled problem for OT-CC, and therefore more straightforward cases of counterbleeding (see Chapter 3 and references therein) will lend themselves to exactly the same analysis.

7.2 Target-like opacity acquisition in OT-CC

In this section we will model the acquisition of CE counterbleeding opacity along the lines of OT-CC (McCarthy 2007). The acquisition model outlined below is fully compatible with BCD (Prince&Tesar 2004) and Minimal GLA (Boersma 2008). Loosely following Mielke, Armstrong & Hume (2003:131), we assume that the following constraints are relevant for our analysis:

(7-3)

Lower V Condition (LVCond)- a lower diphthong [aɪ, ɑʊ] is prohibited before a [-voice] segment in the same foot.

*** V/r[t,d]V** – lax [t] and [d] are prohibited intervocalically or when preceded by /r/ and followed by a vowel.

Higher V Condition (HVCond) – a higher diphthong [əi, ʌʊ] is prohibited.

FaithM: correspondent segments in the input and output have identical manner specifications.

FaithV: correspondent diphthongs in the input and output have identical feature values.

In accordance with BCD (Prince&Tesar 2004) and Minimal GLA (Boersma 2008), we assume that at the initial state markedness constraints dominate faithfulness constraints. On the basis of the evidence presented in Chapter 5, we also assume that the precedence constraint, when constructed, occupies the middle stratum of the hierarchy. The resulting initial-state ranking is shown below:

(7-4) *Ranking at the Initial State*

LVCond, HVCond, *V/r[t,d]V >> Precedence Constraint >> FaithV, FaithM

Following Prince & Tesar (2004), we assume that at the early stages in acquisition the learner goes through the stage of pure phonotactic learning. At this stage the learner is still oblivious to morphology, and therefore such forms as, for example, ‘write’ and ‘writer’ are treated as if they were unrelated to each other. On the basis of the large body of works (see, inter alia, Smolensky 1996a, Kager 1999a, Prince & Tesar 2004), we assume that at this point in acquisition the learner takes lexical representations to be identical to output forms. Therefore, the learner’s task at this stage is to construct a maximally restrictive a grammar that would map each language-legal form to itself. For the limited purposes of our study, we assume that the only forms the learner is exposed to are the ones given below in (7-5). We also assume that the learner is faced with all the forms listed below simultaneously.

(7-5)

[nəif]	‘knife’	[naɪvz]	‘knives’
[ɹəit]	‘write’	[ɹəiɾər]	‘writer’
[ɹaɪd]	‘ride’	[ɹaɪɾər]	‘rider’
[məiɾər]	‘mitre’		

First the learner constructs candidate chains and corresponding candidates for every input, on the basis of her current grammar. These are summarized in Tableau (7-6).

(7-6)

Input	**Candidate Chains**	**Corresponding Candidates**
1. /ɹaɪd/		
	<ɹaɪd> - a faithful parse	(/ɹaɪd/, ɹaɪd, Ø, Ø)
2. /ɹəit/		
	a. <ɹəit> - a faithful parse	(/ɹəit/, ɹəit, Ø, Ø)
	b. <ɹəit, ɹaɪt> - Qd. HVCond >> FaithV	(/ɹəit/, ɹaɪt, FaithV, Ø)
3. /nəif/		
	a. <nəif> - a faithful parse	(/nəif/, nəif, Ø,Ø)
	b. <nəif, naɪf> - Qd. HVCond >> FaithV	(/nəif/, naɪf, Faith V, Ø)
4. /ɹəiɾər/		
	a. <ɹəiɾər> - a faithful parse	(/ɹəɪɾər/, ɹəiɾər, Ø, Ø)
	b. < ɹəɪɾər , ɹaɪɾər > - Qd. HVCond >> FaithV	(/ɹəiɾər/, ɹəɪɾər, Faith V, Ø)
5. /ɹaɪɾər/		
	<ɹaɪɾər> - a faithful parse	(/ɹaɪɾər/, ɹaɪɾər, Ø, Ø)
6. /naɪvz/		
	<naɪvz> - a faithful parse	(/naɪvz/, naɪvz, Ø, Ø)
7. /məiɾər/		
	a. <məiɾər> - faithful parse	(/məɪɾər/, məiɾər, Ø, Ø)
	b. < məiɾər , maɪɾər > - Qd. HVCond>> FaithV	(/məiɾər /, məiɾər, FaithV, Ø)

Having constructed the candidate chains, the learner computes the output of her current grammar in accordance with Minimal GLA (see Boersma 2008). Note, that no Prec constraints are constructed yet, since none of the candidates violates more than one faithfulness constraint. Having done that, the learner constructs loser-winner pairs, in accordance with BCD (see Prince & Tesar 2004).

(7-7)

	LVCond	**HVCond**	*** V/r[t,d]V**	**FaithV**	**FaithM**
1. /ɹəit/					
☞a. ɹəit		✪			
b. ɹaɪt	*			*	
2. /nəif/					
☞a. nəif		✪			
b. naɪf	*			*	
3. /ɹəiɾər/					
☞a. ɹəiɾər		✪			
b. ɹaɪɾər				*	
4. /məiɾər/					
☞a. məiɾər		✪			
b. maɪɾər				*	

(7-8)

Loser < Winner	**LVCond**	**HVCond**	*** V/r[t,d]V**	**FaithV**	**FaithM**
1. /ɹəit/					
b. ɹaɪt < a. ɹəit	W	L		W	
2. /nəif/					
b. naɪf < a. nəif	W	L		W	
3. /ɹəiɾər/					
b. ɹaɪɾər < a. ɹəiɾər		L		W	
4. /məiɾər/					
b. maɪɾər < a. məiɾər		L		W	

In accordance with BCD, the learner first ranks those constraints that prefer neither winners nor losers. Such constraints are the markedness constraint *V/r[t,d]V and the faithfulness constraints FaithM. Since our learner is biased for keeping faithfulness constraints as low as possible, she first ranks *V/r[t,d]V in the upper stratum. Then the learner ranks markedness constraints that prefer only winners, namely, LVCond constraint. Among the constraints left for ranking there is one markedness constraint, namely, HVCond. But this constraint prefers losers over intended winners, and therefore it cannot be ranked yet. In order to free the markedness constraint HVCond for ranking, the learner ranks FaithV in the second stratum, because this constraint prefers only winners. Now there are only two constraints left for ranking. HVCond can be ranked now, because it is safely dominated by FaithV. The remaining faithfulness constraint FaithM goes to the lowest stratum in accordance with BCD. Thus, as a result of pure phonotactic learning, the learner arrives at the following ranking:

(7-9) *Constraint ranking constructed on the basis of pure phonotactic learning*

*V/r[t,d]V >> LVCond >> FaithV >> HVCond >>FaithM

The ranking in (7-9) defines the phonotactics of Canadian English, and therefore no amount of incoming data will be inconsistent with it for as long as the learner holds to the identity map. This ranking is also too permissive, since it allows underlying /əi/ to surface in any position. Therefore, when fed a hypothetical input /kəid/ such grammar would faithfully map it to [kəid]. One way to rule out such a scenario would be to postulate a context-dependent version of HVCond that would explicitly specify that /əi/ is not allowed to surface before voiced segments. Since this constraint is never violated by native forms, by the end of phonotactic learning it would be positioned in the highest stratum together with *V/r[t,d]V. However, we already have one context-sensitive markedness constraint, LVCond. Having a context-sensitive constraint against high diphthongs, therefore, seems redundant. Below we will see that introduction of Prec constraints will help the learner to converge upon the restrictive grammar.

At some point in the acquisition, the learner becomes aware of morphology, and she no longer treats related words in isolation. Below we will model the stage of morphophonemic learning, whereby our child will acquire adult-like underlying representations and refine her phonotactic grammar. We will base our discussion on the algorithm proposed by Tesar & Prince 2003.

As we have said, the learner is now capable of segmenting words into morphemes. She has noticed that certain morphemes alternate depending on their context. The learner's expanded knowledge of morphology is illustrated in the table below.

(7-10) *Morphologically segmented optimal outputs*

Bare Root	**Root + Suffix**
ɹəit$_1$	ɹəiɾ$_1$ + ər
ɹaɪd$_2$	ɹaɪɾ$_2$ + ər
nəif$_3$[18]	naɪv$_3$ + z

---

[18] Regressive voicing assimilation in the mapping /naɪf+z/ → [naɪvz] is an exceptional pattern in English. Normally, the voicing assimilation in plural forms applies progressively, like in /ruf+z/ → [rufs]. Usually, the

The observation that the same morpheme may have different surface forms in different contexts motivates the learner to discard the identity map, and therefore choosing correct underlying representations for alternating morphemes becomes important. According to Tesar & Prince (2003:13), though the learner realizes that assigning a single underlying form that would be identical to both surface instances of an alternating morpheme is impossible, she still adheres to the identity map as closely as possible. In other words, the learner assumes that correct underlying representations of non-varying segments equal their surface representations (further in this section we will present the alternative viewpoint by McCarthy 2004). Then, the learner determines the invariant features of alternating morphemes and fixes them in the underlying form. Having done that, the learner creates several possible underlying representations for each alternating morpheme, differing only in the value of the alternating feature. This is shown in the table below:

(7-11) *Hypothesized underlying forms for alternating morphemes*

Morpheme	**Hypothesized Underlying Forms**
ɹəit$_1$	/ɹəit/$_1$, /ɹəiɾ/$_1$
ɹaɪd$_2$	/ɹaɪd/$_2$, /ɹaɪɾ/$_2$
nəif$_3$	/nəif/$_3$, /naɪf/$_3$

Now the learner has to test her hypothesized lexical representations against the phonotactic ranking. Following Tesar&Prince (2003:15), we assume that at this point the learner does not construct any winner-loser pairs. She just checks if her current grammar renders correct results when fed newly-constructed inputs.

(7-12) *Results for the hypothesized input /nəif/*

	***V/r[t,d]V**	**LVCond**	**FaithV**	**HVCond**	**FaithM**
/nəif/					
☞a. nəif				*	
b. naɪf		*!	*		
/nəif+z/					
💣a. nəivz				*	
☹b. naɪvz			*!		

---

regressive voicing assimilation in plurals is regarded as a lexical exception, like for example in /foot+pl./ → [feet]. For the purposes of this analysis we assume that at early stages of acquisition children treat regressive voicing assimilation as a regular productive process. Exceptional status of the process is discovered later.

(7-13) *Results for the hypothesized input /naɪf/*

	*V/r[t,d]V	LVCond	FaithV	HVCond	FaithM
/naɪf/					
a. nəif			*	*	
b. naɪf		*!			
/naɪf+z/					
a. nəivz			*!	*	
b. naɪvz					

Having tested the hypothesized lexical representations for the morpheme [nəif], the learner has come to the conclusion that only one of them, namely, /naɪf/, produces correct results. Therefore, the learner will store /naɪf/ as the correct underlying representation for the morpheme [nəif].

(7-14) *Results for the hypothesized input /raɪɾ/*

	*V/r[t,d]V	LVCond	FaithV	HVCond	FaithM
/ɹaɪɾ/					
💣a. ɹaɪɾ					
☹b. ɹaɪd					*!
/ɹaɪɾ + ər/					
☞a. ɹaɪɾər					
b. ɹaɪdər	*!				*

(7-15) *Results for the hypothesized input /raɪd/*

	*V/r[t,d]V	LVCond	FaithV	HVCond	FaithM
/ɹaɪd/					
a. ɹaɪɾ					*!
☞b. raɪd					
/ɹaɪd + ər/					
☞a. ɹaɪɾər					
b. ɹaɪdər	*!				*

Having tested the hypothesized lexical representations for the morpheme [ɹaɪd], the learner has come to the conclusion that only one of them, namely, /ɹaɪd/, produces correct results. Therefore, the learner will store /ɹaɪd/ as the correct underlying representation for the morpheme [ɹaɪd].

(7-16) *Results for the hypothesized input /rəiɾ/*

	***V/r[t,d]V**	**LVCond**	**FaithV**	**HVCond**	**FaithM**
/rəiɾ/					
💣a. ɹəiɾ				*	
☹b. rəit				*	*!
/ɹəiɾ + ər/					
☞a. ɹəiɾər					
b. ɹəitər	*!				*

(7-17) *Results for the hypothesized input /rəit/*

	***V/r[t,d]V**	**LVCond**	**FaithV**	**HVCond**	**FaithM**
/ɹəit/					
a. ɹəiɾ				*	*!
☞b. rəit				*	
/ɹəit + ər/					
☞a. ɹəiɾər				*	*
b. ɹəitər	*!			*	

Having tested the hypothesized lexical representations for the morpheme [ɹəit], the learner has come to the conclusion that only one of them, namely, /ɹəit/, produces correct results. Therefore, the learner will store /ɹəit/ as the correct underlying representation for the morpheme [ɹəit]. Or rather 'would have stored' if she were to adhere to the identity map and assume that surface representations of all non-varying segments equal their underlying representation. According to McCarthy 2004, however, the mechanism called the Free Ride, already mentioned in the previous chapters, makes the learner to generalize every unfaithful map discovered from alternations across the entire language. In our case it means that after the learner has discovered that in case of [nəif]-[naɪvz] alternation surface [əi] derives from the underlying /aɪ/, she extends this generalization to non-alternating morphemes and assumes that all instances of surface [əi] derive from underlying /aɪ/. Therefore, she will make another hypothesis, whereby surface [ɹəit] derives from the underlying /ɹaɪt/. She will then test her new hypothesis against her current phonotactic ranking.

(7-18) *Results for the hypothesized input /raɪt/*

	***V/r[t,d]V**	**LVCond**	**FaithV**	**HVCond**	**FaithM**
/ɹaɪt/					
a. ɹaɪt		*!			
☞b. rəit			*	*	
/ɹaɪt + ər/					
☹a. ɹəiɾər			*!	*	*
💣b. ɹaɪɾər					*

The learner realizes that her (only) hypothesis is inconsistent with her current grammar. She attributes this mistake to her current ranking, so she discards the ranking (i.e. she returns to the 'default' initial-state ranking) and constructs winner-loser pairs. In accordance with OT-CC (McCarthy 2007), she constructs candidate chains and corresponding candidates first.

(7-19) *Candidate chains for the input /ɹaɪtər/*

a. <ɹaɪtər> - faithful parse
b. < ɹaɪtər, ɹaɪɾər >, Qd. * V/r[t,d]V >>FaithM
c. < ɹaɪtər, ɹəɪtər >, Qd. LVCond >> FaithV
d. < ɹaɪtər, ɹəɪtər, ɹəɪɾər >, Qd. LVCond >> FaithV and * V/r[t,d]V >>FaithM

(7-20) *Candidates for the input /ɹaɪtər/ ordered as (in, out, L-set, rLUMseq)*

a. (/ɹaɪtər/, ɹaɪtər, Ø, Ø)
b. (/ɹaɪtər/, ɹaɪɾər,{FaithM}, Ø)
c. (/ɹaɪtər /, ɹəitər, {FaithV}, Ø)
d. (/ɹaɪtər/, ɹəɪɾər, {FaithV, FaithM}, <FaithV, FaithM>)

Since one output candidate has an rLUMseq that is not empty, the learner gets evidence to construct precedence constraints, corresponding to both possible sequences of faithfulness violations. Precedence constraints thus constructed are the following:

(7-21)

a. Prec (FaithM, FaithV) – every violation of FaithV must be preceded by a violation of FaithM and must not be followed by a violation of FaithM.
b. Prec (FaithV, FaithM) – every violation of FaithM must be preceded by a violation of FaithV and must not be followed by a violation of FaithV.

The learner constructs winner-loser pairs for the input /ɹaɪtər/. The learner then puts precedence constraints into the grammar one by one to see if they can resolve the conflict. First the learner tests Prec (FaithM, FaithV).

(7-22)

/ɹaɪtər/	**LVCond**	**HVCond**	***V/r[t,d]V**	**Prec (FaithM, FaithV)**	**FaithV**	**FaithM**
a. ɹaɪtər	*		*			
☞d. ɹəɪɾər		✪		✪✪	✪	✪
b. ɹaɪɾər						~~*~~
☞d. ɹəɪɾər		✪		✪✪	✪	~~✪~~
c. ɹəitər		~~*~~	*	~~*~~	~~*~~	
☞d. ɹəɪɾər		~~✪~~		~~✪~~ ✪	~~✪~~	✪

(7-23)

Loser < Winner	**LVCond**	**HVCond**	***V/r[t,d]V**	**Prec (FaithM, FaithV)**	**FaithV**	**FaithM**
a. ɹaɪtər < d. ɹəɪɾər	W	L	W	LL	L	L
b. ɹaɪɾər < d. ɹəɪɾər		L		LL	L	
c. ɹəitər < d. ɹəɪɾər			W	L		L

The learner observes that even with Prec(FaithM, FaithV) in the grammar the intended winner [ɹəɪɾər] is harmonically bounded by the form *[ɹaɪɾər]. The learner tests another constraint, Prec(FaithV, FaithM).

(7-24)

/ɹaɪtər/	LVCond	HVCond	*V/r[t,d]V	Prec (FaithV, FaithM)	FaithV	FaithM
a. ɹaɪtər	*		*			
☞d. ɹəɪɾər		✪			✪	✪
b. ɹaɪɾər				*		~~*~~
☞d. ɹəɪɾər		✪			✪	~~✪~~
c. ɹəitər		~~*~~	*		~~*~~	
☞d. ɹəɪɾər		~~✪~~			~~✪~~	✪

(7-25)

Loser < Winner	LVCond	HVCond	*V/r[t,d]V	Prec (FaithV, FaithM)	FaithV	FaithM
a. ɹaɪtər < d. ɹəɪɾər	W	L	W		L	L
b. ɹaɪɾər < d. ɹəɪɾər		L		W	L	
c. ɹəitər < d. ɹəɪɾər			W			L

The learner observes that with Prec(FaithV, FaithM) it is possible to construct a ranking that would yield correct results. Therefore, Prec(FaithV, FaithM) is stored.

Now the learner updates the set winner-loser pairs she has constructed during phonotactic learning. This is done by means of a process called ‘surgery’ (see Tesar et al. 2003). According to Tesar *et al.* (2003:486) “whenever an underlying form for a morpheme is altered, each winner-loser pair making reference to that morpheme is immediately “adjusted” so that it matches the new underlying form”.

The set of winner-loser pairs now looks as follows:

(7-26)

Loser < Winner	**LVCond**	**HVCond**	***V/r[t,d]V**	**Prec (FaithV, FaithM)**	**FaithV**	**FaithM**
1. /ɹaɪt/						
b. ɹaɪt < a. ɹəit	W	L			L	
2. /naɪf/						
b. naɪf < a. nəif	W	L			L	
5. /ɹaɪtər/						
a. ɹaɪtər < d. ɹəɪɾər	W	L	W		L	L
b. ɹaɪɾər < d. ɹəɪɾər		L		W	L	
c. ɹəitər < d. ɹəɪɾər			W			L

Now the learner applies BCD to the set of winner-loser pairs in order to construct the ranking consistent with the data. Since there are no markedness constraints that prefer neither winners nor losers, the learner ranks two markedness constraints that prefer winners (LVCond and *V/r[t,d]V) in the upper stratum. Among the constraints left for ranking there is one markedness constraint HVCond, but it cannot be ranked at the moment because it prefers some losers. The learner therefore ranks precedence constraint Prec(FaithV, FaithM) in the second stratum, since it only prefers winners. Markedness constraint HVCond is now free for ranking, so it is put into the third stratum. Now the learner has to rank faithfulness constraints. Constraints FaithV and FaithM prefer neither winners nor losers, so they are placed in the bottommost stratum of the hierarchy. The resulting final-state constraint ranking is the following:

(7-27)

LVCond, *V/r[t,d]V >> Prec(FaithV, FaithM) >> HVCond >> FaithV, Faith M

This grammar is such that the violation of HVCond is only permitted in two cases: as a response to LVCond, or as a response to Prec(FaithV, FaithM), which demands that the violation of FaithM be preceded by violation of FaithV. Therefore, underlying /əɪ/ is only allowed to surface before voiceless obstruents. In all other cases it will be mapped to [aɪ].

Thus, we have demonstrated that OT-CC (McCarthy 2007) is capable of accounting for the acquisition of counterbleeding opacity effects with the minimum of machinery and theory-internal stipulation. Moreover, the analysis outlined in this section is consistent with the large body of works on acquisition. Thus, it follows widely held and empirically supported assumptions about the Initial State grammar (Smolensky 1996a, Kager 1999a, Gnanadesikan 1995), it adopts the current solutions to the problem of learning underlying representations (Tesar & Prince 2003, Tesar et al. 2003), and it is fully compatible with BCD (Prince&Tesar 2004) and Minimal GLA (Boersma 2008).

7.3 Target-like opacity acquisition in Stratal OT

There exists at least one alternative to the analysis outlined in Section 7.2, namely, the acquisition model based on Stratal OT (see Bermúdez-Otero 2003). For the sake of comparison, the Stratal-OT-based account will be briefly introduced in this section.

Recall from Chapter 3 that according to Bermúdez-Otero (forthcoming), the main insight behind Stratal OT is that there are intermediate levels of representation between underlying representation and surface representation. Each level of representation is associated with a certain morphological structure. The grammar of a language consists of several strata, each being an OT grammar characterized by a certain constraint hierarchy. Different levels of representation belong to different strata, and the output of one stratum is the input of the next (note that in Stratal OT only the topmost stratum is subject to the Richness of the Base). In the light of Stratal OT, overapplication in [rəiɾər] is analyzed as follows (Bermúdez-Otero, forthcoming: 10):

(7-28) *Overapplication in [rəiɾər] 'writer'*

domain structure: [$_{\mathrm{PL}}$[$_{\mathrm{WL}}$[$_{\mathrm{SL}}$ raɪt] ər]]

Inner cycle (SL)	[rəit] (diphthong raising applies before a voiceless obstruent)
Middle cycle (WL)	[rəitər]
Outer cycle (PL)	[rəiɾər] (flapping applies before a vowel)

Thus, in Stratal OT, opacity is said to arise from the interaction of ordered strata. Note also that at each stratum phonological processes apply transparently, i.e. only when their application criteria are met.

Although Stratal OT has been shown to deal successfully with most attested types of opacity (but see McCarthy 2007 for the discussion on Bedouin Arabic counterfeeding and Wolf 2008 for the discussion on Polish NDEB), some scholars (e.g. Kager 1999a) have raised doubts about the learnability of Stratal OT grammars. However, Bermúdez-Otero (2003) presents an acquisition model that accounts for the acquisition of target-like opacity effects in the light of Stratal OT. For the sake of illustration, the acquisition model as introduced in Bermúdez-Otero (2003) will be briefly outlined below.

According to Bermúdez-Otero (2003), the linguistic data (i.e. adult productions) the child is exposed to are the outputs of the phrase-level co-phonology. Applying phonotactic learning (see Chapter 2 and references therein) to such data, the child can construct the constraint ranking for normally applying surface-true processes (here: flapping). Having done that, the child assigns correct underlying representations to each word in the phrase-level inputs like *hit Ann* vs. *hid Ann*, using evidence from alternations.

Having discovered that surface flap [ɾ] may correspond to either /d/ or /t/, the child quarantines non-alternating word-level items[19] containing flaps: those will be assigned underlying representations later. Next, the child establishes the word-level constraint ranking on the basis of non-quarantined items like *write*, *ride* and *eyeful*. The resulting phonotactic ranking is such that raised diphthongs are allowed when followed by voiceless obstruents, while low diphthongs are allowed in all other contexts. Now the child can lift the quarantine on 'raising' items *writing* and *mitre*, because now she is in the position to assign correct underlying representation /t/ to the flap in these items. Note that items like *powder* and *rider* remain quarantined. Now the learner has to assign the input to the word-level: in doing so, she adheres to the identity map, i.e. [rəitiŋ] for *writing*, etc.

The input to the word-level is also the output of the stem-level. Having removed the word-level suffixes such as *-ful*, what the learner has left is a number of monomorphemic items that obey Raising. On the basis of this evidence, the child constructs the appropriate stem-level ranking. This, in turn, enables the child to assign the correct representations to *powder* and *riding*. Finally, the child should assign lexical representations, i.e. the input to the stem-level

---

[19] Note that according to Bermúdez-Otero (2003:11), at this point the learner treats *writer* and *rider* as monomorphemic.

co-phonology. According to Bermúdez-Otero (2003:14), this is done by analyzing transparent alternations like [səit] - [saɪ'teɪʃn].

Intuitively, the acquisition model proposed by Bermúdez-Otero 2003 might be associated with a number of (potential) problems. Firstly, in order to be able to correctly ascribe morphosyntactic constructions to different strata, learners should be capable of discriminating among such categories as stems, words and phrases at a very early age. Level assignment of morphological units smaller than the word is expected to be especially problematic (see Bermúdez-Otero 1999 on the 'co-phonology arbitrariness problem'; also see Chapter 3 hereof for the discussion). Secondly, the acquisition of multistratal grammar is obviously associated with great computational complexity, especially provided that to date Stratal OT does not have an effective formal mechanism to restrict the differences between the constraint rankings of different co-phonologies (see, *inter alia*, McCarthy 2007 and Kager 1999a for the discussion on this point). Finally, the Stratal-OT-based acquisition model is not straightforwardly compatible with the BCD algorithm (Prince&Tesar 2004). To see if the abovementioned problems can be resolved, further theoretical research and empirical testing are necessary.

7.4 Summary

In this chapter we have seen that acquisition of target-like opacity effects can be modeled in the light of OT-CC (McCarthy 2007) with a minimum of machinery and theory-internal stipulation. In our analysis we have closely followed current stands on such acquisition issues as the nature of the Initial State, acquisition of phonotactics, underlying representations and morphophonology. The acquisition model we have presented is fully compatible with BCD (Prince&Tesar 2004) and Minimal GLA (Boersma 2008). The OT-CC-based acquisition model presented in this chapter can be a good starting point for the further theoretical and empirical research on the problem of acquisition of target-like opacity effects.

Chapter 8: Conclusion

The focus of this thesis is the most recent OT-based theory of opacity called *Optimality Theory with Candidate Chains* (OT-CC, see McCarthy 2007), which represents the synthesis of OT with derivations. We have questioned the status of Precedence constraints and introduced the following changes to the original formulation of OT-CC proposed by McCarthy 2007:

- There is no *B >> Prec (A,B)* metaconstraint on the ranking of Prec constraints.
- Prec constraints are not innate.
- All logically possible Prec constraints are constructed when the learner detects a non-empty rLUMSeq.
- Newly constructed Prec constraints are placed into the hierarchy one by one.
- Prec constraints are subject to the ranking bias, i.e. they are inserted into the hierarchy immediately below markedness constraints and above faithfulness constraints.
- If the learner is trying to acquire an opaque generalization, the Prec constraint that resolves the conflict in the data is kept, all others are discarded.
- If the learner is trying to acquire a transparent generalization, the choice of Prec constraint is incidental and sometimes results in the emergence of developmental opacity effects.

We have demonstrated that the adjustments we have introduced make it possible for OT-CC to account for the emergence and subsequent loss of spontaneous counterbleeding and counterfeeding effects in developing grammars. Importantly, we have shown that though OT-CC predicts (in full consistency with the attested data) that opacity effects might arise in the course of acquisition of transparent generalizations, it does not claim that such effects must and will arise. Thus, our analysis predicts a certain degree of variation in early production data. We demonstrate that constraint rankings corresponding to the attested error patterns arise naturally in the course of acquisition, and the order in which they arise is fully consistent with the attested data. Besides, we have shown that there is a high degree of continuity in transition from one developmental stage to the next. Another obvious advantage of our analysis is that we have avoided the dichotomy problem: the same basic principles and the same formal tools are used to account for both overapplication and underapplication effects in developing grammars. We have also demonstrated that acquisition of target-like opacity effects can be modeled in the light of OT-CC (McCarthy 2007) with the minimum of machinery and theory-internal stipulation. In our analyses we have closely followed current stands on such acquisition issues as the nature of the Initial State, acquisition of phonotactics, underlying representations and morphophonology. Acquisition models we have presented are fully compatible with BCD (Prince&Tesar 2004) and Minimal GLA (Boersma 2008).

It goes without that saying that our assumptions concerning the new status of Prec constraints are largely intuitive, and extensive research is necessary to see if they are empirically and theoretically adequate. Nevertheless, the approach advocated in this thesis has been demonstrated to solve many problems inherent in the previous approaches to opacity, while at the same preserving their merits. Therefore, it makes a good starting point for the further theoretical and empirical research on the problem of acquisition of spontaneous and target-like opacity effects.

Bibliography

Baković, Eric. 2007. “A revised typology of opaque generalizations”. ROA-850

Barlow, Jessica A. and Amanda Keare. 2008. “Acquisition of final voicing: An acoustic and theoretical account”. Indiana University Working Papers in Linguistics: Volume 6.

Becker, Michael. 2006. “CCamelOT – An implementation of OT-CC’s GEN and EVAL in Perl”. Available at [http://people.umass.edu/mbe/papers/becker_ccamelot_lsa.pdf]

Beckman, Jill N. 1997. Positional faithfulness, positional neutralization and Shona vowel harmony. *Phonology* 14:1-46.

Beckman, Jill N. 1998. Positional Faithfulness. ROA-234.

Bermúdez-Otero, Ricardo. 1999. Constraint interaction in language change [Opacity and globality in phonological change.] PhD dissertation. Available at: [www.bermudez-otero.com/PhD.pdf]

Bermúdez-Otero, Ricardo. 2003. “The acquisition of phonological opacity”. ROA-593.

Bermúdez-Otero, Ricardo (forthcoming). Stratal Optimality Theory (Oxford Studies in Theoretical Linguistics). Oxford: Oxford University Press. Partly available at [http://myweb.tiscali.co.uk/bermudez/Stratal_Optimality_Theory.htm]

Boersma, Paul. 1998. “Functional Phonology: formalizing the interactions between articulatory and perceptual drives”.
Available on [http://www.fon.hum.uva.nl/paul/chrono.html]. Last accessed 14.03.2009.

Boersma, Paul. 2008. “Some correct error-driven versions of the Constraint Demotion algorithm”. ROA-980.

Bonilha, Giovana. 2002. “Conjoined constraints and phonological acquisition”. ROA-533

Čavar, Małgorzata. 2005. “Alternating environment in the analysis of derived environment effects”. Available at [https://www.indiana.edu/~iulcwp/pdfs/05-cavar.pdf]

Dinnsen, Daniel A. 2008. “A typology of opacity effects in acquisition”. In Daniel A. Dinnsen and Judith A. Gierut (eds.) *Optimality Theory, Phonological Acquisition and Disorders.* London, Oakville: Equinox

Dinnsen, Daniel A., Laura W. McGarrity, Kathleen M. O’Connor and Kimberly A. B. Swanson. 2000. “On the role of sympathy in acquisition”. Language Acquisition 8:321-361. Available at [http://www.informaworld.com/smpp/content~db=all?content=10.1207/S15327817LA0804_02]

Dinnsen, Daniel A., Kathleen M. O’Connor and Judith A. Gierut. 2001. “The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition”. J. Linguistics 37: 503-525. Available at [http://journals.cambridge.org/download.php?file=%2FLIN%2FLIN37_03%2FS0022226701001062a.pdf&code=4c4b1efbdf1f3da8bcc11dafcdd37ae9].

Dinnsen, Daniel A. and Ashley W. Farris-Trimble. 2008. “An opacity-tolerant conspiracy in phonological acquisition”. Indiana University Working Papers in Linguistics: Volume 6.

Farris-Trimble, Ashley W. 2008. “Cumulative faithfulness effects in phonology”. ROA-991.

Fruehwald, Josef T. 2007. “The Spread of Raising: Opacity, lexicalization, and diffusion”. Available at [http://www.scribd.com/doc/2148281/The-Spread-of-Raising-Opacity-Lexicalization-and-Diffusion]

Gnanadesikan, Amalia. 1995. “Markedness and faithfulness constraints in child phonology”. ROA-67.

Goldrick, Matthew. 2000. “Turbid output representations and the unity of opacity”. ROA-368

Gussmann, Edmund. 2007. *The Phonology of Polish*. Oxford: Oxford University Press

Hayes, Bruce. 2004. “Phonological acquisition in Optimality Theory: the early stages”. In Rene Kager, Joe Pater, Wim Zonneveld (eds.) *Constraints in Phonology Acquisition*. Cambridge: Cambridge University press.

Idsardi, William J. 2000. “Clarifying opacity”. The Linguistic Review 17:337-350. Available at [http://www.ling.udel.edu/idsardi/work/2000tlropacity.pdf]

Idsardi, William J. 2005. “Canadian Raising, Opacity and Rephonemicization”. Available at [http://www.ling.udel.edu/idsardi/work/2005canraising6.pdf]

Jesney, Karen. 2005. Chain shift in phonological acquisition. MA thesis. Available at [http://people.umass.edu/kjesney/papers.html]

Kager, Rene. 1999a. *Optimality Theory*. Cambridge: Cambridge University Press.

Kager, Rene. 1999b. “Surface opacity of metrical structure in Optimality Theory”. ROA-207

Kaye, Jonathan. 2009. “Canadian Raising, eh?” [unpublished manuscript].

Kiparsky, Paul. 1971. “Historical linguistics”. In *A survey of linguistic science*, W. O. Dingwall (ed.), 576-649.

Lubowicz, Anna. 2002. “Derived environment effects in Optimality Theory”. Available at [http://www-rcf.usc.edu/~lubowicz/docs/Lingua.pdf]

Lubowicz, Anna. 2003. “Counter-feeding opacity as a chain-shift effect”. ROA-762

Mielke, Jeff, Mike Armstrong and Elizabeth Hume. 2003. “Looking through opacity”. Theoretical Linguistics. Volume 29, Issue 1-2, Pages 123–139, ISSN (Online) 1613-4060, ISSN (Print) 0301-4428.

McCarthy, John J. 1999. “Sympathy and phonological opacity”. ROA-252

McCarthy, John J. 2000. “Harmonic serialism and harmonic parallelism”. ROA-357

McCarthy, John. 2004. “Taking a free ride in morphophonemic learning”. ROA-683

McCarthy, John J. 2007. *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London, Oakville: Equinox

McCarthy, John J. and Alan Prince. 1993. “Prosodic morphology: constraint interaction and satisfaction”. ROA-482

McCarthy, John J. and Alan Prince. 1995. “Faithfulness and reduplicative identity”. ROA-60

Moreton, Elliott and Paul Smolensky. 2002. “Typological consequences of local constraint conjunction”. ROA-525

Oostendorp, Marc van. 1995. “Vowel quality and phonological projection”. ROA-84

Prince, Alan and Paul Smolensky. 1993. “Optimality Theory: constraint interaction in Generative Grammar”. ROA-537

Prince, Alan and BruceTesar. 2004. “Learning phonotactic distributions”. In Rene Kager, Joe Pater, Wim Zonneveld (eds.) *Constraints in Phonology Acquisition*. Cambridge: Cambridge University press.

Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology: the Structure of Polish*. Dordrecht-Holland/Cinnaminson - USA: Foris Publications.

Sherer, Tim D. 1994. “Prosodic phonotactics”. ROA-54

Smith, Neilson V. 1973. *The Acquisition of Phonology: a Case Study.* Cambridge: Cambridge University Press.

Smolensky, Paul. 1996a.“On the comprehension/production dilemma in child language”. ROA-118.

Smolensky, Paul. 1996b. “The Initial State and ‘Richness of the Base’ in Optimality Theory”. ROA-154.

Smolensky, Paul, Lisa Davidson and Peter Jusczyk. 2004. “The initial and final states: theoretical implications and experimental explorations of the Richness of the Base”. In Rene Kager, Joe Pater, Wim Zonneveld (eds.) *Constraints in Phonological Acquisition*. Cambridge University Press

Stemberger, Joseph Paul, Barbara Handford Bernhardt and Carolyn E. Johnson.1999. "Regressions ("u"- shaped learning) in the acquisition of prosodic structure”. Poster presented at the 6th International Child Language Congress, July 1999. ROA-471

Stemberger, Joseph Paul and Barbara Handford Bernhardt. 2001. “U-shaped learning in language acquisition, and restrictions on error correction”. ROA-472.

Tesar, Bruce. 1995. “Computational Optimality Theory”. ROA-90.

Tesar, Bruce. 2007. “Learnability”. In Paul de Lacy (ed.) *The Cambridge Handbook of Phonology.* Cambridge: Cambridge University Press. Pages 555-574.

Tesar, Bruce and Paul Smolensky. 1993. “The learnability of Optimality Theory: an algorithm and some basic complexity results”. ROA-2.

Tesar, Bruce, John Alderete, Graham Horwood, Nazarré Merchant, Koichi Nishitani and Alan Prince. 2003. “Surgery in Language Learning”. ROA-619.

Tesar, Bruce and Alan Prince. 2003. “Using phonotactics to learn phonological alternations”. ROA-620.

Tessier, Anne-Michelle. 2006. “Stages of phonological acquisition and error-selective Learning”. Available on [http://www.ualberta.ca/~annemich/Papers.html]. Last accessed 14.03.2009.

Vance, Timothy J. 1987. “’Canadian Raising’ in some dialects of the Northern United States”. American Speech. Volume 62, No. 3. Pages 195-210.

Wolf, Matthew. 2008. “Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model”. ROA-996