A theory of morphosyntactic colours

Marc van Oostendorp
Meertens Institute, Amsterdam

July 1, 2006

Contents

1 Introduction: Faithfulness with and without OT 2
  1.1 A preliminary definition .......................... 2
  1.2 Containment and Invariance ...................... 4
  1.3 Lexicon Optimization .............................. 7
  1.4 The Correspondence Relation and Constraints .... 11
  1.5 Attacks on Consistency of Exponence ............ 13
  1.6 Special faithfulness violations: Metathesis ....... 19
  1.7 Faithfulness in reduplication .................... 24
  1.8 Distinctions between faithfulness and markedness .... 26
  1.9 Lexical Phonology ................................ 30
  1.10 Arguments for and against Lexical Phonology .... 32
  1.11 Coloured Containment ............................ 37
  1.12 Morphological colouring .......................... 39
  1.13 Opacity and output representations .............. 44

2 Across morphological boundaries 47
  2.1 Asymmetries between prefixes and suffixes .......... 47
  2.2 Lexical levels ...................................... 51
  2.3 Inflectional suffixes ............................... 56

3 Minimal morpheme expression 59
  3.1 Introduction ........................................ 59
  3.2 Nasal assimilation in Hellendoorn Dutch past tense .... 63
  3.3 Voicing Assimilation in Flanders and Brabant ....... 68
  3.4 Remnants of the first person singular morpheme .... 72
Chapter 1

Introduction: Faithfulness with and without OT

1.1 A preliminary definition

The term ‘faithfulness’ has been introduced into phonology by Prince & Smolensky (1993) and McCarthy & Prince (1993), and it has become one of the cornerstones of Optimality-Theoretic analysis. It is usually assumed within Optimality Theory (OT) that grammars consist of two types of constraints: markedness constraints, requiring the output to have some ‘optimal’ shape and faithfulness constraints which require the outputs not to differ unnecessarily from the inputs. Faithfulness constraints keep the markedness constraints in balance; even though it can also happen that markedness constraints conflict among themselves, the most common type of conflict is one between a markedness constraint and a faithfulness constraint. For instance if the input is /CVC/, the following three constraints are relevant:

(1) a. NOCODA: Syllables should not end in a consonant.

1.1 A preliminary definition
b. PARSE-C: Do not delete consonants.
c. FILL-V: Do not insert vowels.

NoCoda is an obvious markedness constraint, requiring outputs to have open syllables only. If it were on its own in the phonological universe, all languages would do anything to avoid consonants. Inputs /CVC/ would be turned into outputs [CVCV] or [CV] everywhere.

OT exists because constraints like NoCoda are not alone in the phonological universe. The faithfulness constraints PARSE-C and FILL-V serve two functions:

1. They are used in the description of languages in which words do not obey the universal markedness constraint.
2. They are used to describe which strategy is chosen if a language does obey the markedness constraint and underlying representations do not conform to it.

On the one hand faithfulness constraints make sure that NoCoda is counterbalanced in some languages, so that they have coda’s. In this case we have PARSE-C, FILL-V $\gg$ NoCoda:

\[
\begin{array}{|c|c|c|c|}
\hline
/tat/ & PARSE-C & FILL-V & NOCODA \\
\hline
\hline
ta & *! & & \\
\hline
\hat{t}a.tat & & * & \\
\hline
ta.ta & *! & & \\
\hline
\end{array}
\]

The second function is to decide which strategy is chosen in a language which does not allow coda’s: deletion of onsets or deletion of coda’s. This strategy is famously dependent on the relative ordering of the faithfulness constraints which are involved. One ordering favours CVC $\rightarrow$ CV, the other CVC $\rightarrow$ CVCV:

\[
\begin{array}{|c|c|c|c|}
\hline
/tat/ & NOCODA & FILL-V & PARSE-C \\
\hline
\hline
\hat{t}a & & * & \\
\hline
tat & *! & & *! \\
\hline
ta.ta & & *! & \\
\hline
\end{array}
\]

a.

\[
\begin{array}{|c|c|c|c|}
\hline
/tat/ & NOCODA & PARSE-C & FILL-V \\
\hline
\hline
ta & & *! & \\
\hline
tat & *! & & \\
\hline
\hat{t}a.ta & & & * \\
\hline
\end{array}
\]

The examples I have just given are standard examples of the way in which OT works, and you can find them in any textbook on this topic (Kager 1999).
People therefore tend to think that faithfulness is a notion that is specific for OT. But in a sense the notion has been an important one for generative phonology during all of its history. The issue arises in every theory which has more than one level of representation, and in which a mapping has to occur.

**Definition 1 (Faithfulness)** Suppose \( R_1, R_2 \) are linguistic representations, which are linguistically related. Faithfulness requirements are requirements that force \( R_1 \) and \( R_2 \) to be identical.

This definition is very imprecise; it depends on two notions which are themselves in need of definition: *linguistically related* and *identical*. With respect to the former, we will assume for now that this is defined by a derivational relationship (one form is derived from the other); furthermore, in this class we will only consider pairs of representations where \( R_1 \) is an underlying form, i.e. a lexical representation and \( R_2 \) is the output of phonology, i.e. a phonetic representation. We will see in later chapters that there probably are no researchers who restrict faithfulness to uniquely this situation; but it nevertheless constitutes the core of the theory.

### 1.2 Containment and Invariance

The notion ‘faithfulness’ was first introduced explicitly in the literature by [Prince & Smolensky (1993)](https://www.jstor.org/stable/2433894). One of the central axioms on which this classical version of OT was based was *Containment*:

**Definition 2 (Containment)** Every element of the phonological input representation is contained in the output. (There is no deletion.)

Containment conflicted heavily with another aspect of Classical OT, which actually was a far more radical departure of previous approaches, Richness of the Base:

**Definition 3 (Richness of the Base)** There are no conditions on underlying representations. Anything can be input to the grammar of language and still give a grammatical output.

While Containment gives us some very basic (but powerful) form of faithfulness, Richness of the Base runs against all faithfulness. Since inputs can be anything, including wildly ungrammatical forms — a series of three hundred clicks for English —, in some cases the phonetic output has to be completely different from the input. But how can we get from one to the other, if at the same time we are not allowed to delete anything?
The answer of Containment Theory (CT, this is how we will refer to the theory outlined in [Prince & Smolensky](1993) from now on) is that the mapping from underlying representation to phonetics runs in two stages:

1. In the first stage, we have an OT system, not technically deleting elements which will not make it to the surface but marking them as somehow unpronouncable; the ‘output’ of this stage is still a fairly abstract representation.
2. In the second stage, we map the output of stage 1 to the phonetics. The way this mapping would work has never been worked out (as far as I am aware), but at least it should involve the deletion of the elements which were marked as unpronouncable before. Otherwise it should be fairly uninteresting and language-independent.

The system as a whole does not satisfy Containment, but the first stage does, and this is where the interest of the phonologist is supposed to be — the place where the important generalisations are. The output of the phonology thus is quite remote from phonetics. A priori this does not seem to be either a good thing or a bad thing, but something that could be verified by empirical study (even though this may be very hard).

The way deletable elements are marked in CT is by ‘not parsing’ them. Segments that should not be pronounced are not incorporated into syllable structure, for instance, and features are not parsed into segments, i.e. not connected to them by association lines. It is clear that none of these options were available in SPE phonology for which syllable structure and autosegmental tiers did not exist; this is the reason why the results of [Chomsky & Halle](1968) about Invariance do not carry over immediately to the new framework.

Faithfulness does not just prohibit deletion of material; it also prevents insertion. Again, the approach to this in CT was fairly abstract. One could insert material, but at the output of the phonology one could still see that it was inserted. Epenthetic vowels somehow looked differently from underlying vowels, and similarly for consonants. Constraints against insertion were constraints against segments that looked different. The way this was implemented was by making epenthetic segments completely empty, i.e. devoid of any feature content. The features were then filled in at the second stage of the phonology (the ‘phonetics’).

This way of implementing deletion and insertion make CT into a purely monostratal theory; constraints do not need to refer to any phonological level of representation beyond the output. Also faithfulness constraints refer to this level — and there thus is no formal difference between faithfulness and markedness constraints. Unfaithful outputs are unwellformed outputs, since they contain unparsed or empty material.

The input thus was not just contained in the output, but it could actually be reconstructed from it. Consider the following phonological output repre-
sentation for phonetic [takp]:

\[
\begin{array}{c}
\sigma & \sigma \\
\text{t a k p} & \emptyset
\end{array}
\]

From this, we know that the input must have been /takp/. The final vowel is empty — possibly pronounced as [ə] —, hence it must be epenthetic; the /k/ is not incorporated into syllable structure; hence it will not be pronounced.

One obvious problem with this implementation of the principle of Containment is that it uses aspects of representation that have been used to describe phenomena of purely faithful phonology as well. It has been argued that extraprosodic segments can stay outside syllable structure, and floating features have been part and parcel of many autosegmental analyses. However, we may still assume that extraprosodic segments are still associated to some higher-level structure (to the Phonological Word rather than the syllable), and it is not necessarily a problem that floating features are deleted in the phonetics, after they have enforced their phonological effect. The more serious problem lies in the representation of epenthetic material: it cannot be denied that epenthetic vowels and consonants often are the target of spreading, but this cannot be represented if they have to remain empty in order to be recognizable as potential violators of faithfulness constraints.

On the other hand, CT does not face all of the problems of Invariance. As we have noted above, the key reason for this is that our view of representations has changed. The key example for Chomsky & Halle (1968) was the fact that the final vowel of algebra is [-tense] if the word occurs in isolation, and [+tense] if the word occurs before the suffix -ic. We thus need to change a feature value. But if we assume that the relevant feature is monovalent [(+)lax] (van Oostendorp, 2000), there is no problem in analysing this in a way which satisfies Containment, viz. by letting [lax] float (it was always assumed that Containment did not apply to association lines):

\[
\begin{array}{c}
\text{a l g e b r a (i c)} \\
\text{[lax] [lax] [lax]}
\end{array}
\]

The second component will delete the floating feature [lax] and at the same time, mark all vowels which do not have such a feature as [tense].

Insertion of features again would be more problematic, since there is no device to mark inserted material as violators. We therefore cannot allow fea-

\footnote{A feature [lax] is associated to the second vowel here only to make the analysis more similar to the one in SPE. The vowel /i/ probably also is lax, but this is not pictured here for the sake of simplicity.}
ture insertion at all – if it would be allowed, there would be no way to block random insertion of features, except for markednes. Furthermore, we would also be able to insert features into epenthetic segments, thus freeing them from their status of empty segments. Thus, Chomsky’s Problem is not solved in a satisfactory way.

This also implies that the input representations of OT have to be rather different from those of SPE also in another respect: the heavily underspecified matrices of (??) could only be filled in at the second stage, and therefore anything interesting that could be said about them would fall beyond the scope of phonology proper. This basically means that we cannot have underspecification, as we will see in section 1.3. Thus, the underlying forms will have to be quite close to the output representations. This is true also in another respect. In this sense the faithfulness requirements may be too strong.

The main advantage of a CT approach to faithfulness is that it is parsimonous: it does not refer to any device which may not be needed independently. For instance, the PARSE and FILL families of constraint supposedly are necessary beyond the theory of faithfulness. This poses a problem for many of the proposed alternatives to be discussed in following chapters: we may stipulate that we no longer use PARSE-C or FILL-V, but then we will still need to say something about consonants that are not attached to syllable nodes on the surface or vowels that do not have any vowel content. In that sense, CT comes very close to the null hypothesis regarding faithfulness theory, given the other theoretical assumptions that were made in Prince & Smolensky (1993), McCarthy & Prince (1993).

1.3 Lexicon Optimization

The fact that CT is based on an assumption of Richness of the Base, does not imply that it does not recognize the existence of a lexicon, where the precise linguistic structure of individual items is stored, of course. It is clear that every language has a lexicon, and it is also clear that for instance a child acquiring a language should learn the arbitrary form-meaning pairs that are found in the lexicon. What Richness of the Base does say, however, is that nothing interesting can be said about the structure of the lexicon. All generalisations (e.g. the fact that there are no English words starting with *rt-) are epiphenomena from the constraints which operate on the output of the phonology (there is a constraint there forbidding *rt-, so that words of this shape will never survive).

How do we select the shape of the lexical items? Although it is usually assumed that every input corresponds to exactly one output (so-called free variation aside), the inverse is not true. For instance, we know that English does not have the sound [O]. Suppose we also know that underlying /O/ is deleted (rather than turned into, say, [t]). This implies then that there are
1.3. Lexicon Optimization

infinitely many possible inputs for the word [man]: /man, òman, òòman, òòòman, . . ./ How does the child select the ‘real’ underlying form? Prince & Smolensky (1993) suggest that in this case a reverse optimisation takes place. We do not optimize the surface form, but the underlying representation; interestingly, this can be done using the same constraint ranking as for ‘normal’ output optimization. The following is Prince & Smolensky’s first attempt:

**Definition 4 (Lexicon Optimization)** Suppose that several different inputs $I_1, I_2, \ldots, I_n$, when parsed by a grammar $G$ lead to corresponding outputs $O_1, O_2, \ldots, O_n$, all of which are realized as the same phonetic form $\Phi$ — these inputs are all phonetically equivalent with respect to $G$. Now one of these outputs must be the most harmonic, by virtue of incurring the least significant violation marks: suppose this optimal one is labelled $O_k$. Then the learner should choose, as the underlying form for $\Phi$, the input $I_k$.

It is easy to see what ‘the most harmonic input’ will be like, approximately. All inputs lead to the same output by definition, and therefore they will all incur exactly the same number of markedness violations. Hence the faithfulness constraints will decide, and they will pick the input which will give the smallest number of faithfulness violations. This will be the input that is closest to the output. And given Richness of the Base, there will be an input which is exactly identical to the output, so this will be selected. Lexical representations can thus be expected to be exactly the same as their pronounced form.

An obvious question arises at this point: how should we treat cases where segments alternate in different phonological contexts? To take one example out of many — any textbook on phonology is obviously full of them — consider Root Final Devoicing of stops in Turkish. The question is discussed by Prince & Smolensky (1993), but here we will rather follow the line of reasoning of Inkelas (1995). Consider the following data:

(6) a. Alternating root-final plosive:
   kanat ‘wing’ kanad-ı ‘wing-Acc’
   kanat-lar ‘wing-pl’ kanad-im ‘wing-1sg.poss’

b. Nonalternating voiceless plosive:
   sanat ‘art’ sanat-ı ‘art-Acc’
   sanat-lar ‘art-pl’ sanat-im ‘art-1sg.poss’

The problem is with the examples in (6a): the procedure for Lexicon Optimization cannot tell us about the voicing quality of the final plosive in the underlying representation (since it cannot satisfy faithfulness both with the nominative singular and the accusative singular). Therefore, Inkelas (1995) proposes an alternative definition:
1.3. Lexicon Optimization

Definition 5 (Alternation-sensitive restatement of L.O.) Given a grammar $G$ and a set $S = \{S_1, S_2, \ldots, S_i\}$ of surface phonetic forms for a morpheme $M$, suppose that there is a set of inputs $I = \{I_1, I_2, \ldots, I_j\}$, each of whose members has a set of surface realizations equivalent to $S$. There is some $I_i \in I$ such that the mapping between $I_i$ and the members of $S$ is the most harmonic with respect to $G$, i.e. incurs the fewest marks for the highest ranked constraints. The learner should choose $I_i$ as the underlying representation for $M$.

It is not specified how the calculation works, exactly, or what exactly are the members of $S$ (if two suffixes trigger the same form of the stem, should this form be counted once or twice), but all of this could of course be worked out in a satisfactory way.

It is interesting to see (although this was not noted by Inkelas) that this Alternation-sensitive L.O. effectively derives Kiparsky (1982a)'s Alternation Condition at least in the sense that underlying forms with feature values which conflict with all surface instances of the same morpheme are not allowed. (Recall that Kiparsky did not consider underspecification as an option.)

For the Turkish cases, we will now get the following Lexicon Optimization tableaux, where $D$ denotes a coronal stop that is not specified for voicing:

\[
\begin{array}{|c|c|}
\hline
\text{L.O.} & \text{PARSE- [+voice]} \\
\hline
\text{a. /kanad/} & \text{[kanat]} & \text{*!} \\
& [kanadi] & \\
\hline
\text{b. /kanaD/} & \text{[kanat]} & \\
& [kanadi] & \\
\hline
\end{array}
\]

In order to evaluate Lexicon Optimization, we thus need to evaluate several related forms (both with and without affixes) at the same time; as a matter of fact we need to compare all the environments in which a morpheme can occur. This involves something beyond the normal comparison of individual forms; rather we need some form of output-output faithfulness, as will be discussed in chapter ??.

The relevant faithfulness constraint is a PARSE-constraint against deletion of features, presumably since there is no ban on inserting features in CT, as we have seen above. If a segment is underspecified, no feature has to be deleted. Note that it is actually essential that there is no constraint against insertion of features, since otherwise it would not be so clear which of these input forms would win. And note also that it is equally crucial that underspecified ‘archisegments’ such as /D/ should not be allowed to surface as such. This is the case of all underspecification theories, and it means that a force of antifaithfulness is as work as well: underlying representations have properties which surface representations do not have.
Turkish actually provides good empirical evidence for underlying underspecification, since it has a three-way contrast. Next to words which show alternation, and forms which are uniformly voiceless, there are also (loan)words which are uniformly voiced:

(8) Nonalternating voiced plosive:

\[ \text{etüd} \quad \text{etüd-i} \quad \text{etüd-ı} \quad \text{etüd-ım} \]

We can analyse this by assuming that \text{sanat} ends underlyingly in \([-\text{voice}], \text{etüd} \text{ in } [+\text{voice}] \text{ and } \text{kanat} \text{ in } [0\text{ voice}]. \text{ We can give these words the following analysis (FINDEV is a stand-in for whatever is responsible for Final Devoicing):}

(9)

\begin{align*}
\text{a.} & \quad /\text{etüd}/ & \text{PARSE}[-+\text{voice}] & \text{FINDEV} \\
& \text{etüd} & * & \\
& \text{etüd} & *! & \\
\text{b.} & \quad /\text{sanat}/ & \text{PARSE}[-+\text{voice}] & \text{FINDEV} \\
& \text{sanat} & & \\
& \text{sanad} & *! & \\
\text{c.} & \quad /\text{kanat}/ & \text{PARSE}[-+\text{voice}] & \text{FINDEV} \\
& \text{kanat} & & \\
& \text{kanad} & *! & \\
\end{align*}

Notice that the ranking \text{PARSE}[-+\text{voice}] \gg \text{FINDEV} is crucial to get the right result for \text{etüd}. In a language like German, which does not have exceptions of this kind, we need a ranking \text{FINDEV} \gg \text{PARSE}[-+\text{voice}]. Lexicon Optimization then will mean in turn that there will be no words in German with an underlyingly voiced plosive, as the reader may verify.

Languages like English are somewhat problematic, however. The fact that English does not have final devoicing effects means that it has the ranking \text{PARSE}[-+\text{voice}] \gg \text{FINDEV}, just like Turkish. But the question then is: why doesn’t English have alternating forms of the type \text{kanat}? We cannot simply stipulate that English does not have the relevant archiphonemes, since this would violate Richness of the Base; but it is unclear how the problem should be solved in any other way.

The problem becomes even more complicated if we look at the analysis of the Turkish cases with a vowel-initial suffix. We have to make sure that the unspecified forms take the voiced alternative in this case (\text{kanadı}, *\text{kanatı}). In order to do this, we may postulate a constraint causing intervocalic voicing; let us call it IV. A constraint \text{PARSE}[-\text{voice}] must dominate it, in order to make sure that the plosive in \text{sanat} does not also voice:
1.4 The Correspondence Relation and Constraints

It is no exaggeration to say that Correspondence Theory is the standard theory of faithfulness within Optimality Theory. In this chapter, we will discuss the standard version of the theory, as it was presented in McCarthy & Prince (1995a). Several extensions to Correspondence Theory will be discussed in the chapters to follow. Here, we will concentrate on the standard version of the theory.

The main motivation for replacing Containment with Correspondence was that some kind of faithfulness relation seems to hold between the two parts of a reduplicated form. Consider, for instance, the following examples from Javanese:

\[
\begin{array}{|c|c|c|}
\hline
\text{i. Stem} & \text{ii. +C} & \text{iii. +V} \\
\hline
\text{a. anEh} & \text{anEh-ku} & \text{anE-e} \\
\text{b. b@da} & \text{b@da-b@da} & \text{b@da-b@da-e} \\
\text{c. dajO} & \text{dajO-dajO} & \text{dajO-dajO-e} \\
\hline
\end{array}
\]

Again, the question arises why other languages which do not have Final Devoicing, also do not have the pattern here.

We can thus conclude that Lexicon Optimisation is not without its problems; as far as I am aware most of these problems have not been addressed, let alone answered in the literature. Later developments in faithfulness theories can be shown also not to really shed light on these matters.

1.4 The Correspondence Relation and Constraints

It is no exaggeration to say that Correspondence Theory is the standard theory of faithfulness within Optimality Theory. In this chapter, we will discuss the standard version of the theory, as it was presented in McCarthy & Prince (1995a). Several extensions to Correspondence Theory will be discussed in the chapters to follow. Here, we will concentrate on the standard version of the theory.

The main motivation for replacing Containment with Correspondence was that some kind of faithfulness relation seems to hold between the two parts of a reduplicated form. Consider, for instance, the following examples from Javanese:

\[
\begin{array}{|c|c|c|}
\hline
\text{i. Stem} & \text{ii. +C} & \text{iii. +V} \\
\hline
\text{a. anEh} & \text{anEh-ku} & \text{anE-e} \\
\text{b. b@da} & \text{b@da-b@da} & \text{b@da-b@da-e} \\
\text{c. dajO} & \text{dajO-dajO} & \text{dajO-dajO-e} \\
\hline
\end{array}
\]

Again, the question arises why other languages which do not have Final Devoicing, also do not have the pattern here.

We can thus conclude that Lexicon Optimisation is not without its problems; as far as I am aware most of these problems have not been addressed, let alone answered in the literature. Later developments in faithfulness theories can be shown also not to really shed light on these matters.
though the first half itself does not occur before a vowel and thus does not satisfy the relevant context. The reason for this seems to be that the ‘reduplicant’ needs to be as similar to the ‘base’ as possible: in other words a relation of faithfulness seems to hold between them.

Obviously, this type faithfulness cannot be expressed in terms of CT, since this requires absolute identity of the faithful objects (they are literally the same thing), yet the two [b]’s of [bəda-bəda-e] clearly represent different phonetic events, and similarly for the other segments. The faithfulness requirements of Correspondence Theory are more relaxed: it does not demand absolute identity, but merely that the two segments stand in a relation to one another. This relation is called ‘correspondence’:

**Definition 6 (Correspondence)** Given two strings $S_1$ and $S_2$, correspondence is a relation $\mathcal{R}$ from the elements of $S_1$ to those of $S_2$. Elements $\alpha \in S_1$ and $\beta \in S_2$ are referred to as correspondents of one another when $\alpha \mathcal{R} \beta$.

It should be noted that here, as in the formulation of Containment we have seen, a special role in the transmission of faithfulness is awarded to the phonological segment (rather than features or syllables): the relation is defined in terms of strings, and in the practice of Correspondence Theory this has always remained to be the focus of attention. This is in line with the rather minimalist assumptions regarding representations that most OT practitioners follow.

The most important innovation of Correspondence Theory, however, is that also input-output pairs no longer have to be identical. Indeed, a segment in the input can be completely distinct from a segment in the output. This gives us a potential way of solving Chomsky’s problem: we may assume in principle that the input representation is a purely cognitive object, whereas the output is a completely phonetic object. As long as we can establish a relation $\alpha \mathcal{R} \beta$, nothing forces us to assume that $\alpha$ and $\beta$ have to be qualitatively the same thing (although obviously they are the same types of object in reduplication).

This also means that we can no longer see that something is inserted or deleted if we only look at the output. Instead of the singular representation that appeared in the tableaux of Containment Theory (4) the following pair of representations, plus the correspondence relation between them, now has to be evaluated by the grammar:

<table>
<thead>
<tr>
<th>(12)</th>
<th>Input</th>
<th>Output</th>
<th>Correspondence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t\alpha$ aβ k, pδ</td>
<td>$\sigma$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td></td>
<td>$t\alpha$ aβ</td>
<td>$p_d \circ e$</td>
<td>$\mathcal{C}(a_\beta, a_b)$</td>
</tr>
<tr>
<td></td>
<td>$t\alpha$ aβ</td>
<td>$p_d \circ e$</td>
<td>$\mathcal{C}(p_b, p_d)$</td>
</tr>
</tbody>
</table>
1.5 Attacks on Consistency of Exponence

We know that /kγ/ has been deleted, because this segment is present in the underlying representation, and there is no segment x in the output such that C(kγ, x). Similarly, [œ] has been inserted because it is in the output, and there is no segment y in the input such that C(y, œ).

The downside of this is that the objects on which our computations are based, have become quite complex. In OT tableaux using Correspondence Theory, most scholars still draw only the output representations, and this can be justified from a presentational point of view since the input is the same for all forms in a tableau. Yet this does not represent the real computational complexity of the problem involved. In reality, all candidates have a structure which is as complicated as (??). It is also fairly obvious that these representations are not pronounced as such, so that we still need a minimal rule component which takes out the ‘output’ representation and at least delete the indices which are necessary for the computation of correspondence relations.

In other words, correspondence has it that the analysis of every word in natural language involves reduplication, but where the base is silent. The two representations S1 and S2 have to be maximally similar because of a number of faithfulness constraints such as the following:

(13) a. The MAX Constraint Family
    Every segment of S1 has a correspondent in S2.

b. The DEP Constraint Family
    Every segment of S2 has a correspondent in S2.

c. The IDENT Constraint Family
    Let α be a segment in S1 and β be any correspondent of α in S2. If α is [γF], then β is [γF]. (Correspondent segments are identical in feature F).

These are constraint families rather than constraints, because there are different versions of them for input-output (‘IO’) correspondence and for base-reduplicant (‘BR’) correspondence (these have to be separate because they can have independent positions in the hierarchy (for instance BR faithfulness has to be ranked higher than IO faithfulness, since similarity to the base can enforce dissimilarity to the input).

1.5 Attacks on Consistency of Exponence

Even though Consistency of Exponence is usually ignored in the literature, the only explicit attacks against it of which I am aware are [Walker & Feng (2004)] and [Łubowicz (2005a)]. These authors suggest that Consistency of Exponence is not a restriction of Gen, but a set of violable constraints in CON, controlling the output of Gen. Gen will be able to change the morphological affiliation of phonological material, but in many cases these constraints
will filter out candidates with those changes. The argumentation in favour of such a position should be that Consistency of Exponence is sometimes violated by a winner candidate in some natural language; this is the type of argument Walker & Feng and Łubowicz try to provide.

The data are from Anxiang, a Chinese dialect spoken in the Hunan Province in central China. In this dialect, a diminutive is formed of a noun by adding Car, where C is a copy of the stem consonant, and ə an epenthetic vowel.

<table>
<thead>
<tr>
<th>Stem</th>
<th>Diminutive</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>pʰᵃᵖʰᵃᵖʰᵃᵖʳ</td>
<td>pʰᵃᵖʰᵃᵖʰᵃᵖʳ</td>
<td>‘claw’</td>
</tr>
<tr>
<td>kᵉᵏᵉʳᵃʳ</td>
<td>kᵉᵏᵉʳᵃʳ</td>
<td>‘square’</td>
</tr>
<tr>
<td>tᵒᵗᵒʳᵃʳ</td>
<td>tᵒᵗᵒʳᵃʳ</td>
<td>‘pile’</td>
</tr>
<tr>
<td>pʰʷᵘ⁻ʷᵘᵖʰʷᵘᵖʰʷᵘᵖʰʷᵘᵖʰʳ</td>
<td>pʰʷᵘ⁻ʷᵘᵖʰʷᵘᵖʰʷᵘᵖʰʷᵘᵖʰʳ</td>
<td>‘shop’</td>
</tr>
<tr>
<td>pʰᵃᵘ⁻ᵃᵘᵖʰᵃᵘᵖʰᵃᵘᵖʰʳ</td>
<td>pʰᵃᵘ⁻ᵃᵘᵖʰᵃᵘᵖʰᵃᵘᵖʰʳ</td>
<td>‘bulb’</td>
</tr>
</tbody>
</table>

The diminutive form thus consists of the stem, a copy of the consonant of the stem, a schwa and an /r/. Since the schwa may be considered an epenthetic vowel, Walker & Feng assume that the underlying structure of the diminutive suffix is /r/.

Walker & Feng (2004) argue that we need two constraints for our analysis of these facts. First we have the following constraint on the morphology-phonology interface:

(15) \text{ALIGN}: ‘Each morpheme should occupy exactly one syllable and vice versa’

This constraint is argued to be a very general one on the phonology-morphology interface in this dialect of Chinese in which every morpheme has the size of exactly one syllable — and every syllable therefore corresponds to exactly one morpheme.

Next to this well-formedness constraint, Walker & Feng (2004) propose the following faithfulness constraint, which is essentially a violable version of Consistency of Exponence:

(16) \text{IDENT-MM}: ‘Let \(\alpha\) be a morpheme in the input, and \(\beta\) be its correspondent morpheme in the output. If \(\alpha\) has phonological content \(\phi\), then \(\beta\) has phonological content \(\phi\) and vice versa.’

In order to show that IDENT-MM is not a restriction on Gen, we need to show that it is sometimes violated in the winning candidate. This is exactly what happens in Anxiang, according to Walker & Feng (2004), as illustrated by the following tableau:
The winning candidate here is the one in which the ‘copied’ consonant and the epenthetic vowel have become part of the morpheme on the surface. This would prove that Consistency of Exponence is indeed violable.

Yet very much depends on our interpretation of the constraint $\text{ALIGN}[\sigma]$. Under one interpretation, we could argue that this constraint is violated in (17a) both by the morpheme $ke$- and by the morpheme $-r$ — because neither corresponds exactly to one syllable —, so that this form has 2 violations of that constraint. On the other hand, (17b) has only 1 (for $r$, but not for $ke$). This means that even if we do not take (17c) into consideration — because this is not generated under Consistency of Exponence — the correct surface string $kek\sigma$ would still win:

Walker (p.c.) suggests a slightly different interpretation:

“The $\text{ALIGN}[\sigma]$ constraint is actually a cover constraint for four alignment constraints operating over syllables and morpheme edges [i.e. $\text{ALIGN}(\mu, L, \sigma, L)$, $\text{ALIGN}(\mu, R, \sigma, R)$, $\text{ALIGN}(\sigma, L, \mu, L)$ and $\text{ALIGN}(\sigma, R, \mu, R)$, where $\sigma=a$ syllable and $\mu=a$ morheme] [...] $\text{ALIGN}(\mu, L, \sigma, L)$ and 1 wrt $\text{ALIGN}(\mu, R, \sigma, R)$. (17b) incurs 1 wrt $\text{ALIGN}(\mu, L, \sigma, L)$ and 1 wrt $\text{ALIGN}(\sigma, L, \mu, L)$. However, (17c) obeys each of the four alignment constraints under our interpretation whereby it violates Consistency of Exponence.

If the four constraints that compose $\text{ALIGN}[\sigma]$ are tallied together, violation of Consistency of Exponence is crucial. But this implies that a ranking of $\text{ALIGN}(\mu, R, \sigma, R) \gg \text{ALIGN}(\sigma, L, \mu, L)$ might obtain the desired result.”

I take this to mean that the Anxiang example as yet does not provide conclusive evidence that against an inviolable interpretation of Consistency of Exponence.\footnote{Walker & Feng (2004) note that “Violation of Consistency of Exponence has also been observed in Turkish (Karabay, to appear) and in the Jian’ou Chinese dialect (Feng, in prep.).” This evidence is as yet inaccessible, and can therefore not be evaluated.}
Although it is a little less explicit, another proposal for a violable interpretation of Consistency of Exponence can be found in [Łubowicz (2005a)], who analyses phenomena in Palauan and Akkadian, and argues that these should be understood as the result of ‘morpheme absorption’: infixed elements become part of the morphological stem. I will discuss the Palauan facts here, but I believe a similar reanalysis can be made for Akkadian.

In Palauan, then, there is a morpheme /m (a)/ which behaves sometimes as a prefix and sometimes as an infix; the status of the schwa is irrelevant. [Łubowicz (2005a)] argues that the choice between these two options is non-phonological and made on morphological grounds only:

(19) Prefixation | dakt ‘fear’ m-o-dakt ‘be/get fearful’
                 | rur ‘shame’ m-o-rur ‘be/get ashamed’
Infixation      | l-aŋol ‘crying’ l-m-aŋol ‘cry’
                 | rurt ‘running’ r-o-m-urt ‘run’

The prefix and the infix behave differently with respect to one phonological phenomenon: (long distance) dissimilation. If the verb already contains a labial consonant, the infix nasal turns into a rounded vowel [u], whereas the prefix nasal is not affected:

(20) Prefixation | dub ‘poison’ m-o-dub ‘be/get poisoned/bombed’
                 | kimud ‘cut hair’ m-o-rur ‘been cut (hair)’
Infixation       | r-ebat ‘action of falling’ r-u-ebat ‘fall (from)’
                 | ?-árm ‘suffering’ ?-u-áram ‘suffer’

[Łubowicz (2005a)] claims that the relevant constraint is a version of the OCP indexed for morphological category:

(21) OCP<sub>root</sub>(C-lab): Avoid more than one labial consonant in the root domain.

The idea is that this constraint is violated in the infixed cases, but not in the prefixed cases. But this idea only makes sense if the infix is part of the root — it has been ‘absorbed’ by it —, whereas the prefix stays outside.

In order to get this effect, [Łubowicz (2005a)] invokes two constraints: one is a violable version of Consistency of Exponence, which she dubs MORPHEME-DEPENDENCE (M-DEP), and another constraint MORPHEME-LOCALITY (M-LOC) which disallows discontinuous morphemes:

(22) a. M-DEP: Let M<sub>i</sub> be a morpheme and S<sub>j</sub> be a phonological element in two related morpho-phonological representations, M<sub>i</sub> and S<sub>j</sub> ∈ Input, M<sub>i</sub> and S<sub>j</sub> ∈ Output, M<sub>i</sub> ▽ M<sub>2</sub> and S<sub>j</sub> ▽ S<sub>2</sub>, If S<sub>j</sub> ∈ M<sub>i</sub>, then S<sub>j</sub> ∈ M<sub>i</sub>. 
1.5. Attacks on Consistency of Exponence

b. M-LOC: Let M be a morpheme, and xyz be segments, where xyz ∈ Output: If xyz are adjacent, and x ∈ M ∧ z ∈ M, then y ∈ M

A ranking M-LOC >> M-DEP gives the right morphological parsing (given that the status of the morpheme as an infix is predetermined by the morphology):

(23)  a. | /l-m-atk/ | M-LOC | M-DEP |
     | [lmatk]   |   *  |
     | [lm]atk   |   *! |

b. | /m-dakt/ | M-LOC | M-DEP |
   | [m]dakt   |   *!  |

In (23a), the morphology has decided that /m/ is an infix; however, M-LOC does not like to see infixes at the surface, and it therefore turns the segment into a part of the stem, creating a violation of M-DEP. On the other hand, there is no potential problem with M-LOC in (23b), so that in this case M-DEP (=Consistency of Exponence) decides.

This analysis suffers from a number of problems. In the first place, it has to determine the placement of affixes as infix or prefix before any phonology can take place. The reason for this is that in the output form [lmatk], the morphology would no longer be able to check that /m/ has indeed turned into an infix: as a matter of fact the relevant adjectivising suffix has become completely invisible, since it is ‘absorbed’ by the stem. This means that morphological constraints can no longer be operative at this level. This implies a type of serialism however — doing morphology before phonology — which runs counter to one central tenet of classical OT, where the placement of affixes can be determined by the interaction of phonological and morphological constraints (Prince & Smolensky (1993); see also Golston (1995)).

Given this architecture of the grammar, it is furthermore unclear why we need to have morphological structure in the output of the phonological module at all. The morphological constraints are no longer operative at this level, so why would we need to see the difference between stems and affixes at all? The answer to this is: because phonological constraints like OCP,root,C-lab need to see them. But this actually turns the root in these cases into a purely prosodic category, which is only there to establish the domain of phonological phenomena.

If that is the case, however, we might just as well abandon the problematic assumption that morphological grammar precedes phonological grammar, and reformulate the OCP constraint directly into prosodic terms:
(24) OCP\textsubscript{PW}(C-lab): Avoid more than one labial consonant in the phonological word.

In this case, we will have to make sure that infixes become part of the phonological word, whereas prefixes stay outside. We do not need to relativize Consistency of Exponence for this, however. Under the assumption that embedded phonological word structures such as (25a) are universally impossible, we only need a constraint for Alignment of phonological and morphological categories. And even if we would allow Gen to produce structures such as this, we could still have a high-ranking constraint against such center-embedding phonological structures, paralleling Lubowicz\textsuperscript{(2005a)}’s M-LOC.

\begin{center}
\begin{tikzpicture}
\node (phi) at (0,0) {$\phi$};
\node (phi) at (0,-1) {$\phi$};
\node at (1,-2) {\textsuperscript{a}lm a tk};
\end{tikzpicture}
\end{center}

(25) a. *lm a tk

b. Φ-LOC: Phonological words must be contiguous (=M-LOC, applied to phonological words).

c. ALIGN: The edges of a phonological word should correspond to the edges of a morpheme.

We can now build exact parallels in (23):

(26) a. \[
\begin{array}{c|c|c}
\text{latk}/+/	ext{m} & \Phi-\text{LOC} & \text{ALIGN} \\
\hline
\varepsilon\phi(lmatk) & \ast & \\
(l(m)atk) & \ast! & \\
\end{array}
\]

b. \[
\begin{array}{c|c|c}
\text{dakt}/+/	ext{m} & \Phi-\text{LOC} & \text{ALIGN} \\
\hline
\varepsilon\phi(m\text{d}akt) & \ast! & \\
(m\text{d}akt) & \ast! & \\
\end{array}
\]

I believe that this analysis is not just a notational variant of the one which turns Consistency of Exponence into a violable constraint; it is conceptually superior to it on the grounds of interleaving morphological and phonological constraints in a way which is well-established and of not using morphological categories as phonological diacritics. Instead of this, it uses the category of the phonological word, which is independently needed. We just see that neither Walker & Feng\textsuperscript{(2004)}’s nor Lubowicz\textsuperscript{(2005a)}’s attempt to turn Consistency of Exponence into a violable constraint is very successful. From this we conclude that its status as a restriction on Gen is uncontested.

MAX-constraints ban deletion of segments (everything in the input needs to be in the output); DEP-constraints militate against insertion (everything in
the output needs to be in the input). Both constraints only demand that there is some segment in the other level of representation, and do not care about the quality of the other segment. They are satisfied if underlying a corresponds to phonetic [t]. The latter configuration is prohibited by IDENT-constraints, since many of the features of /a/ are not found on its correspondent [t].

Notice that, just as in the case of Containment Theory, nothing in this formulation prohibits filling in underspecified features: if the underlying representation (S1) does not have a feature value for [F], IDENT as it is formulated is vacuously satisfied. This also means that this formulation does not work satisfactorily with theories of unary features (cf. Blaho, 2004). The present formulation furthermore assumes that the underlying representation and the output representation are the same (so we do not work with phonological vs. phonetic features), but obviously a slight reformulation could be made to work under contrary assumptions.

1.6 Special faithfulness violations: Metathesis

The constraints mentioned above form the core of the faithfulness constraints. There also is a number of other constraints which regulate other types of faithfulness. Some of these could be formalized in Containment theory if we wanted to:

(i) a. The I-CONTIGUITY Constraint Family
   The portion of S1 standing in correspondence forms a contiguous string. (No deletion of elements internal to the input string.)
   b. The O-CONTIGUITY Constraint Family
   The portion of S1 standing in correspondence forms a contiguous string. (No deletion of elements internal to the input string.)

In order to express this, we could prohibit empty and unparsed segments in the middle of a string. It is not obvious that this would be a more or a less plausible interpretation of the phenomenon at hand than what we have in Correspondence Theory. Something similar could be said about the Anchoring constraints:

(ii) a. The I-ANCHORING-(L/R) Constraint Family
   The portion of S1 standing in correspondence forms occurs on the left/right edge string. (No deletion of elements at the left/right to the input string.)
   b. The O-ANCHORING-(L/R) Constraint Family
   The portion of S2 standing in correspondence forms occurs on the left/right edge string. (No insertion of elements at the left/right to the input string.)
We return briefly to this issue in section 1.12. Others however are impossible to formulate under Containment:

(27)  a. The **Uniformity** Constraint Family

“No Coalescence” No element of $S_2$ has multiple correspondents in $S_1$.

b. The **Integrity** Constraint Family

“No Breaking” No element of $S_1$ has multiple correspondents in $S_2$.

c. The **Linearity** Constraint Family

“No Metathesis”: $S_1$ is consistent with the precedence structure of $S_2$, and vice versa.

It is clear that **Linearity** could not be expressed within Containment Theory; there is no way to mark that two segments have changed their order. Similarly, we cannot tell from looking at the output alone whether two underlying segments are reversed or, inversely one underlying segment has been split up into two, so that **Uniformity** and **Integrity** also are impossible to express.

It is not immediately uncontroversial, however, that this should be counted as an advantage for Correspondence Theory. Clearly, it makes the theory much less restrictive than Containment. In the former theory, faithfulness was partly built into Gen, and the core of it was universal. Within Correspondence theory, any input can correspond to any output, hence all faithfulness is violable. This is a good thing if we see OT purely as a descriptive instrument which should be able to handle any phenomenon we encounter without any problem. Yet, from the point of view of explanatory adequacy, the theory does not fare very good at all. It is very hard to conceive logically possible but inattested phenomena that could be modelled perfectly in this theory of containment. For instance, suppose we have a language L, which has stems /ta/, /bo/, /at/, and two suffixes -/k/, and /i/. Suppose furthermore that the syllable structure constraints **NoHiatus** and **NoCluster** are very high ranked in this language, and so are all faithfulness constraints, except for (**IO** **Linearity**). We then get as a result that words metathesize according to the suffix:

(28)  a. 

<table>
<thead>
<tr>
<th>/ta/+/k/</th>
<th>NOCLUSTER</th>
<th>FAITH</th>
<th>LINEARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOHiatus</td>
<td></td>
<td></td>
</tr>
<tr>
<td><code>w[tak]</code></td>
<td></td>
<td>#1</td>
<td></td>
</tr>
<tr>
<td><code>[atk]</code></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
1.6. Special faithfulness violations: Metathesis

Underlying /ta/ and /at/ thus give the same outputs. It is not difficult to see that this result applies also to longer words, so that /takepi/ and /atekip/ are also indistinguishable. But languages of this type, where the order of consonants and vowels can be freely reversed according to the context, have never been attested. The problem is that the three constraint families which are difficult to handle for containment (Uniformity, Integrity, Linearity) do not seem as freely rankable as other constraints are. They are always lowly ranked. Even though the processes of metathesis, fusion and fission are well-known from historical phonology, and to some extent maybe also from reduplication, it is not clear that we cannot do without them in the synchronic phonology. In the absence of strong evidence in its favour (and such evidence can only consist of solidly analysed data, not just some random examples from languages which have not been carefully studied) there is no reason to assume the more permissive theory.

Interestingly, McCarthy (2000) gives precisely such an argument. He discusses the so-called ‘incomplete phase’ in Rotuman, a central Oceanic language spoken on an island about 500 kilometres north of Fiji (Churchward, 1940). The incomplete phase is, roughly, an environment in which the last two syllables of a word turn into one syllable. There are several ways of achieving this (not all of them listed here), but one of them is metathesis:

(29) Phase Differences in Rotuman
1.6. Special faithfulness violations: Metathesis

<table>
<thead>
<tr>
<th></th>
<th>Complete</th>
<th>Incomplete</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Deletion</td>
<td>tokiri</td>
<td>tokir</td>
</tr>
<tr>
<td></td>
<td>ti?u</td>
<td>ti?</td>
</tr>
<tr>
<td></td>
<td>'big'</td>
<td></td>
</tr>
<tr>
<td>b. Metathesis</td>
<td>i?q</td>
<td>ia?</td>
</tr>
<tr>
<td></td>
<td>seseva</td>
<td>seseav</td>
</tr>
<tr>
<td></td>
<td>'fish'</td>
<td>'erroneous'</td>
</tr>
<tr>
<td>c. Diphthongization</td>
<td>pupui</td>
<td>pupũi</td>
</tr>
<tr>
<td></td>
<td>lelei</td>
<td>lelei</td>
</tr>
<tr>
<td></td>
<td>'floor'</td>
<td>'good'</td>
</tr>
</tbody>
</table>

McCarthy (2000) shows that in case the two vowels are separated by a consonant, metathesis is preferred over vowel deletion. The latter option is only chosen if the former would result in a non-rising diphthong. Clearly, this implies that we somehow need to incorporate a device of metathesis into our theory, and McCarthy claims that this indeed provides us with an argument against Containment Theory:

If input /pure/ must be “contained” in every candidate, then literal puer is not even in the candidate set, since /pure/ \( \not\subset \) puer, stringwise. Under Containment, the real output form should be puEr<e>, with epenthetic E copying the unparsed final vowel. Yet how can this form, which violates both FILL and PARSE, ever compete with *pur<e>*, which violates only PARSE? To put the matter differently, if metathesis is copy+deletion, how will it ever be favored over straight deletion, since deletion receives a subset of the marks that metathesis receives? To escape this consequence of its core assumptions, the PARSE/FILL model would need to introduce considerable elaboration of the conception of what is PARSEd or FILLed, perhaps distinguishing effects on the Root nodes of the vowels from their Place nodes.

The latter remark (probably) means that we could save Containment Theory if we would be allowed to have independent faithfulness of Place nodes: the Place node of the deleted vowel could then be moved to the epenthetic vowel to preserve Place faithfulness. It is hard to decide whether the ‘elaboration’ this requires is really more ‘considerable’ than introducing the complicated machinery of metathesis into our theory, in particular since it seems clear that natural languages have processes of vowel copying, by which the full content of a vowel spreads across a consonant to another (epenthetic) vowel; the most well known example of this is Yawelmani.

McCarthy (2000) also mentions another possible analysis of this phenomenon, which would equally mean a more sophisticated theory of representations, viz. one which is based on ‘tier segregation’ — the idea that vowels and consonants occur on separate tiers underlyingly, and are not (totally) ordered on that level of representation with respect to each other, so that the incomplete phase does not really constitute a violation of faithfulness. McCarthy is highly critical of this analysis as well:
With tier segregation, the possibility of metathesis is hard-wired into every phonological representation — in essence, it’s a stipulated regularity of the Rotuman lexicon which has overt consequences in the incomplete phase. In Optimality Theory, on the other hand, there are no stipulated regularities in the lexicon; regularities must emerge from the constraints on output forms, possibly augmented with assumptions about language learning (Prince & Smolensky 1993: Chapt. 9). There are several reasons for pursuing this claim in OT: it is presupposed by another claim, that all interlinguistic differences are to be derived from constraint ranking; it is possible to obtain the effect of constraints on underlying representation from output constraints; and it is necessary to do so, in order to solve the “duplication problem” (on which see Kenstowicz & Kisseberth 1977). […] More broadly, OT takes much of the burden of explanation off of representations (e.g., tier segregation) and places it on substantive constraints (e.g., LINEARITY), which are violable under domination. Thus, consonant-vowel tier segregation is completely superfluous in an Optimality Theoretic analysis of Rotuman, and in fact it is antithetical to fundamental premises of OT.

This argument is based on the implicit assumption that tier segregation is not available for languages which do not display its effect. If it would be a universal possibility in human language, it would not be a ‘stipulated regularity of the Rotuman lexicon’, nor would we need to abandon the claim that all language variation can be derived from constraint ranking. Furthermore, the duplication problem also would not arise, since this problem concerns only language-specific characteristics (which are duplicated since they are stated both in the lexicon and in the grammar).

Related to this is that this discussion is based on a specific interpretation of Richness of the Base, viz. one where we assume that nothing can be said about inputs at all. Another interpretation is different, however, viz. that we cannot use underlying representations as a way of marking systematic differences between languages. Under such an interpretation, Optimality Theory is a theory which mainly serves to describe (interlinguistic and intralinguistic) variation. Universals could be made to follow from a theory of representations, i.e. (partly) a theory of possible underlying representations.

Under such a theory, it is not clear at all, whether we need LINEARITY. This becomes even more apparent if we consider the following concession McCarthy makes:

While V-C metathesis is not uncommon, to my knowledge processes of V-V metathesis have been proposed on just three occasions […]. All three involve very abstract analyses, in which the underlying representations and/or the consequences of metathesis are by no means apparent, and all except Latvian have been
reanalyzed in ways that do not involve V-V metathesis at all. It therefore seems reasonable to prohibit V-V metathesis outright, perhaps universally [...].

If we want to distinguish between V-C metathesis on the one hand, and V-V metathesis on the other, we need to take recourse to some theory of representations. We could wonder whether this could then not take care of the other aspects of metathesis as well.

More in general, we need very strong arguments in order to convince us that we should make the step from Containment Theory to Correspondence Theory. The strongest argument seems to be the fact that reduplication obeys the same patterns. If it can be shown that we need the correspondence machinery independently to describe reduplication, there may be no real extra ‘cost’ to adapting it for input-output relations as well.

1.7 Faithfulness in reduplication

Let us now see how correspondence relations behave in reduplication. We have already seen one example above, from Javanese, in which an /h/ had disappeared on the surface in the reduplicant only because it also disappeared from the base (example (11) on p. 11). This is traditionally called ‘overapplication’, since it would be described in rule-based terms by a rule (/h deletion/) which would apply in a form which does not meet its condition. Another example is from Tagalog. In this language there is a process of ‘nasal substitution’. After a prefix ending in a velar nasal, a stem initial plosive turns into a nasal (e.g. /p/-[m]).

\[(30)\]
\[
a. \text{putul ‘cut (n.)’} \\
b. \text{pang-putul ‘that used for cutting’ >}[\text{pamutul}] \\
c. \text{pa-mu-mutul ‘a cutting in quantity’ (reduplication; *pa-mu-putul)}
\]

Nasal substitution overapplies in the reduplicated form: the first consonant in the base is not preceded by a velar nasal, but still it turns into a nasal. The reason for this apparently is that in this way it is more similar to the reduplicant.

Overapplication is a good argument for faithfulness within reduplicated forms, since there seems to be no other reason why it would happen. Another good argument is its reverse, underapplication: cases where some phonological rule does not seem to apply even though its phonological condition is met. An example for this may come from Japanese (as has been observed in Kager (1999), based on data from Mester & Itô (1989)). In Tokyo Japanese, [g] and [i] are famously in complementary distribution. The former occurs at the beginning of the word, whereas the latter occurs in other positions:
Complementary (allophonic) distributions like this in general pose some puzzles for faithfulness theories (cf. section ??); but one way of seeing this certainly is that that /g/ turns into [ŋ] in the middle of the word. Assuming that we have a constraint against velar nasals in word-initial position (*[wN]) and another constraint against velar voiced stops (*g), and assuming that all the relevant IO faithfulness constraints are low ranked, we get the right result:

(32) a. /geta/ or /Neta/ | *[wN] *g

   | *w

   | *!

b. /kagi/ or /kaNi/ | *[wN] *g

   | *w

   | *

Now this process does not apply in so-called ‘mimetic’ reduplication:

(33) a. gara-gara ‘rattle’ (*garaNara, *NaraNara)
b. geji-geji ‘centipede’ (*gejiNeji, *NejiNeji)
c. gera-gera ‘laughing’ (*geraNera, *NeraNera)

‘Normal’ application of the nasalisation process would lead to *garaNara, etc. Overapplication would lead to *NaraNara, etc. The form we find, however, are cases of underapplication: the process does not apply, even though it could in principle. BR Faithfulness is able to describe this:

(34) /gara/+red or /Nara/+red | BR FAITH | *[wN] | *g

   | *w

   | *

   | *!

   | *!

Obviously, this is dependent on many other assumptions we make. For instance, if we assume that the two parts of the reduplication both constitute a domain D which is relevant to the constraint *[Dŋ], we do not need to refer to BR Faithfulness at all:
1.8. Distinctions between faithfulness and markedness

An important question now is whether the Japanese example just happens to be an unfortunate one, or whether other purported examples would also be subject to alternative analyses (we will return to a different type of reduplication in Tokyo Japanese in sections ?? and [1.10]). There is at least one attempt at giving an alternative to the BR Faithfulness account within OT: this is [Inkelas & Zoll (2005)]. They discuss the Javanese case mentioned above; the analysis is fairly complicated, but the general idea is that the reduplication is not a prefix, as would be assumed under a BR Faithfulness analysis, but rather an infix. This means that there is some level at which the stem is next to the prefix:

(36) putul ↓ + pang
    pang-mutul ↓ + RED
    pang-mu-mutul

The way the analysis is represented here, shows that this analysis requires a faithfulness theory for the morphology-phonology interface which is more complicated than a simple input consisting of a stem plus a set of suffixes. We need to incorporate a notion of cyclicity, i.e. we need faithfulness between several levels which are defined by the addition of one single affix at a time. (We will return to the issue of cyclicity in section ??.) The Javanese example presented above could be analysed in a similar way. (See [Keenan (2002); Frampton (2001)] for a formal criticism of Correspondence Theory as a theory of reduplication and [Raimy (2000); Zoll (2002); Frampton (2004)] for empirical arguments and alternative analyses.)

We thus have to conclude that we need some type of faithfulness to understand reduplication, and that for this we cannot use the simplest form of faithfulness theory. It is not so clear, however, that it means we have to abandon Containment. In the model presented in (36), we could use Containment in principle.

1.8 Distinctions between faithfulness and markedness

There is yet another potential advantage of Correspondence Theory over Containment Theory. We have seen that Containment Theory used markedness constraints (against empty and unparsed segments) to express faithful-
1.8. Distinctions between faithfulness and markedness

ness. On the other hand, Correspondence Theory uses a separate formalism for faithfulness constraints. (To be precise, McCarthy & Prince (1995a) suggest that Correspondence could be used also for autosegmental association, and for ALIGNMENT, but neither of these suggestions have been taken up in full detail.)

Dividing up the set of constraints formally in this way has certain implications, and these actually seem to be very favourable for Correspondence Theory. The reason is that certain subtheories actually refer to the difference between faithfulness and markedness.

The most well-known example of this is language acquisition. It is a standard assumption in the OT acquisition literature that in the initial stage all children have a grammar in which all markedness constraints outrank all faithfulness constraints (cf. the contributions to Kager et al. 2004, for an overview). One argument for this is that in this way we can explain the so-called Jakobsonian Generalisation (Jakobson, 1942; Smolensky, 1996):

**Definition 7 (Jakobson’s Generalisation)** Structures which are avoided via phonological ‘processes’ within adult languages, and excluded from some inventories across adult languages, also tend to be structures which are later-acquired by children.

The avoidance or exclusion of structures in adult languages can only be accomplished by markedness constraints dominating faithfulness in OT. If we assume that this is the initial state, Jakobson’s Generalisation follows trivially.

Another argument for this assumption is that it allows for a solution of the so-called Subset Problem of language acquisition. Smolensky (1996) gives the following example of this. Suppose we have a language L which has CV Syllables only (so all words are CV, CVCV, CVCVCV, ...). Every grammar (i.e. every permutation of the set of universal constraints) can generate words with CV syllables. The question arises: which grammar does the language learner acquire when she is confronted with the evidence of language L? A priori it is in no position to decide that closed syllables should be excluded from the language, given the lack of negative evidence. In other words, it does not know whether it is learning the \{CV\}∗ language L, or some \{CVC\}∗ language Lc. (The former language is a subset of the latter, this is why this is called the Subset Problem.) Yet all the available evidence seems to show that children who learn a language such as L, will assume that closed syllables are disallowed: the subset language is assumed, unless we find (positive) evidence that we need to assume the superset.

Within OT, there is an obvious way of handling this, viz. by assuming that initially the markedness constraint NOCODA dominates the faithfulness constraints MAX-C and DEP-V. More generally, we may assume, once again, that the initial state is one in which all faithfulness constraints are outranked by all markedness constraints. This will give a grammar which has precisely one output for all inputs, viz. the most unmarked form of the language (say
[\text{ta}] or [ti] or something similar). Only if we find positive evidence that something more marked is possible, we can rerank one markedness constraint and one faithfulness constraint to get the desired result.

We thus see that two problems of language acquisition (the Jakobsonian Generalisation and the Subset Problem) can be solved at the same time by the same assumption, viz. that initially markedness dominates faithfulness. But this assumption only makes sense if we can really distinguish the two types of constraints. The distinction between faithfulness and markedness constraints also plays a role in the analysis of other phonological phenomena as well. \text{Itô \& Mester} (1995, 2001) have argued that we can use this distinction also to understand the ‘stratal’ organisation of the lexicon in natural language. In many languages, there is a distinction between native words and loanwords, for instance. Roughly, loanwords are more exceptional than native words. We can understand this in the following way: for loanwords some faithfulness constraints are higher-ranked than for native words.

\text{Itô \& Mester} study languages such as Japanese and Korean in which the structure of the lexicon is more refined than a mere distinction between loanwords and native words. Their theory of lexical strata is roughly as follows. All the markedness constraints in a given language have a fixed ranking \( M_1 \gg M_2 \gg \ldots \gg M_n \). The faithfulness constraints can be reordered with respect to this fixed ranking. In the more ‘native’ parts of the lexicon, we will have lower-ranked faithfulness, in the more ‘foreign’ (or learned) part of the lexicon, we will have higher faithfulness.

An example from Jamaican Creole (Meade, 2001; \text{Itô \& Mester}, 2001) may illustrate this point. This is an English-based creole in which several phonological processes apply optionally to acrolectal forms. Thus we have a process of cluster simplification (e.g., [st]ick → [t]ick) and a process of hardening of voiced interdentals (e.g., [ð]at → [d]at). Acrolectal [ðat stik] (that stick) thus appears as basilectal [dat tik]. Interestingly, we also find the mesolectal form pronunciation [dat stik] (with hardening but without simplification) but the fourth logical possibility *[ðat tık] (i.e., with simplification but without hardening) is not attested.

\text{Meade} (2001) and \text{Itô \& Mester} (2001) view this as evidence that there is a fixed ranking of the markedness constraints (37a), and ranking faithfulness relative to this gives us exactly the three attested varieties (37c):

\begin{enumerate}
\item \text{\texttt{[ð]at stik}}
\item \text{\texttt{[dat tik]}}
\end{enumerate}
1.8. Distinctions between faithfulness and markedness

The only way in which we could get the unattested form in (37b) would be by reranking the markedness constraints, but apparently this option is not available to us. We thus need to be able to refer to the difference between markedness and faithfulness constraints in order to describe this.

There is a third way in which the distinction between the two types of constraints is relevant, but this is very close to what we have just seen. In van Oostendorp (1997) it is argued that differences between styles of speech (i.e. between more formal and more informal speech levels) within a language can be understood by reference to this distinction. E.g., in Dutch, we can distinguish between at least three styles of speech with respect to the behaviour of vowel reduction. In the most formal variety the word *phonology* is pronounced as [fɔnɔlɔˈɡi] without reduction; in a slightly less formal variety we find [fɔnɔlɔˈɡi], with the posttonic vowel reduced; in the most informal variety we find [fɔnɔlɔˈɡi], with all unstressed vowels reduced. Unattested is the form [fɔnɔlɔˈɡi], with only the pretonic unstressed vowel reduced. We can understand this, again, by assuming there is a fixed ranking of markedness constraints, and that faithfulness constraints can be reranked with respect to these constraints according to the (possibly universal) principle in

(38) a. The more formal the style of speech, the higher ranked faithfulness constraints are.

b. ← Faith-C (formal)   no reduction [fonoloˈyi]

*POSTTONICFULL
| ← Faith-B (informal)   reduction of post-tonic [fonoloˈyi]

*UNSTRESSEDFULL
| ← Faith-A (very informal)   full reduction [fonɔlɔˈyi]

The principle in (38a) also applies to the Jamaican Creole case, if we assume that the acrolect is the more ‘formal’ variety and the basilect the more ‘informal’. This shows the conceptual similarity between the difference between (learned) loanwords vs. native words on the one hand, and formal vs. infor-
mal speech on the other. These differences can both be described in terms of the relative ranking of faithfulness vis à vis markedness.

In the ideal case, this would be reflected in a formal distinction between the two types of constraints. Correspondence Theory allows us to do this:

- **Faithfulness constraints** refer to correspondence relations between input and output.\(^3\)
- **Markedness constraints** refer to output representations.

Even though it was customary to refer to faithfulness vs. markedness constraints already in the Containment literature, it is not clear how we could make a sharp formal distinction between the two types of constraints there. We will see later on, however, that the distinction sometimes becomes blurred in further developments of Correspondence Theory as well.

### 1.9 Lexical Phonology

There was some derivational residue already in one of the first manuscripts on OT, McCarthy & Prince (1993): the use of different phonological levels, as they were introduced in the framework of Lexical Phonology (LP Kiparsky 1982b, 1983, 1985, 2000; Mohanan 1986; Hargus & Kaisse 1993; Booij 1997). Within this theory, the phonological component is divided into at least two subcomponents: lexical and postlexical phonology. These are seen as independent phonological systems, one functioning on words, and the other on linguistic expressions which are bigger than words. Within OT, each of the components consists of a Generator and an Evaluator function, built on a separate constraint ranking. They are ordered with respect to one another: the output of the lexical phonology functions as the input to the postlexical phonology. In this sense, the theory is derivational.

It is very common to also distinguish between different levels within the lexical phonology. In their analysis of Axininca Campa, McCarthy & Prince (1993) distinguish between three different levels:

⇒ Prefix Level ⇒ Suffix Level ⇒ Word Level

A more common distinction is the following:

⇒ Stem Level ⇒ Word Level ⇒ Post-Lexical Level

It is an important empirical question of LP whether all languages have the same level structure. Another question is whether a language would allow random rerankings of constraints between different phonological levels.

\(^3\)Strictly speaking, this makes only MAX and DEP constraints into real faithfulness constraints within Correspondence Theory, because only they can be understood by reference to the existence of correspondence relations alone. In order to check, e.g. IDENT or LINEARITY constraints, we also need to look at the content of the input and output representations. We will ignore this technical problem.
Both questions as yet await a definitive answer. But a more pressing question is whether we need the apparatus of Lexical Phonology, given the richness of the theory of faithfulness already provided by Correspondence Theory.

This question is in the first place an empirical one. A strong argument in favour of LP is that in many languages we find a much more restricted syllable structure at the lexical level than at the phonetic level. For instance, it makes sense to say that French onsets are maximally bipositional and satisfy the usual sonority requirements. Yet on the surface, we find many structures such as the following:

(39) *Henri devrait partir* ‘H. would have to leave’ [āridvrēpartir]

We would like to say that the cluster [dvr] corresponds to two onsets at some level of the phonology. We could state that this is a matter of OO Correspondence (to the isolation form [dvr]), but this begs the question why isolation forms are always more restricted than forms which do not occur in isolation. Booij (1997) points out that

“what the derivational metaphor of evaluation in two steps expresses is that the postlexical phonology may make the effects of the lexical phonology opaque, whereas the inverse, lexical phonology making the postlexical phonology opaque, does not occur.”

He provides an example of this type of opacity from Dutch. As we have already seen above, this language has syllable-final devoicing. This property becomes opaque however in encliticized structures:

(40) *(ik) heb het* ‘I have it’ /heb at/ [he.bɔt]

*(hij) had het* ‘he had it’ /had at/ [ha.tɔt]

*(ik) heb ’r* ‘I have her’ /heb ər/ [he.bɔr]

This is an instance of opacity because the context for syllable-final devoicing is absent (the obstruents are in an onset), but the process still applies. Of course, a Sympathy or OO analysis might be feasible, but they would not explain why it is that small domains influence big domains rather than the other way around. The relation is not symmetric: it could be symbolized by an arrow. But arrows are what the derivational ‘metaphor’ is all about: as soon as we have an arrow in a relation between representations, we could see the relation as derivation; other interpretations might be possible as well, but it then is the question whether those other interpretations have different empirical ramifications, i.e. whether such a revised OO Faithfulness account would not be a notational variant of Lexical Phonology.

Furthermore, Booij (1997) points out that an OO-approach would need different faithfulness constraint rankings for cliticised forms and for suffixes. The latter do not show the effects of final devoicing (e.g. *hebben* ‘to have’ [he.bɔn]). In a two-level approach, we can do with one grammar which is
applied twice, and in which the coda devoicing constraint outranks faithfulness, as you can check for yourself. Still, certain changes will need to be made from one structure to another; interestingly, Booij (1997) tries to restrict these changes to resyllabification, in accordance with (but independent from) Faithfulness-Free Syllabification (def. ?? on p. ??).

We can therefore reverse the question we put at the beginning. Given the fact that Sympathy Theory and OO Faithfulness seem to explain the same facts as Lexical Phonology, we could wonder whether the former should not be done away with. This in turn would imply a severe blow for Correspondence Theory, which seems to imply these notions; it would be an argument in favour of a more restricted version of faithfulness theory, e.g. Containment. In the next section, we will see how proponents of the combination of LP and OT have solved some of the puzzles we have seen above, as well as a case which has been put forward as an argument in favour of OO Faithfulness rather than LP.

1.10 Arguments for and against Lexical Phonology

We return first to the facts of Tokyo Japanese discussed in section ???. We have seen there that the opaque behaviour of nasalisation in Rendaku is hard to describe, given the fact that it involves allophonic variation, and this cannot be captured with faithfulness, since we have to assume that the underlying representation is underdetermined. We have also seen that the issue would not be problematic anymore if we could assume (as classical accounts do) that the input is in fact always /g/, never /ŋ/. The way this can be accomplished in Lexical Phonology is as follows: we could first make sure that the lexical component turns every /ŋ/ into /g/; and then the postlexical component could change intervocalic initial /g/ into /ŋ/ again. (Or alternatively, we could turn every voiced velar into /ŋ/ lexically, and then turn initial /ŋ/ into /g/ in the postlexical component.)

We can see immediately that this makes the theory somewhat less restrictive than Cumulativity, since it allows the DY effect /ŋ/ → /g/ → /ŋ/. But it seems that we have no other choice for cases such as these, where we have to fix an input for some form of allophonic variation. It might be, then, that in exactly in cases like this we do not some limited form of DY power. If we want to keep to Richness of the Base, on the one hand, and assume that certain forms on allophonic variation are introduced ‘later’ than other decisions (regardless of whether we do this with sympathy, OO Faithfulness or lexical phonology), this is the conclusion we should draw.

The intuition behind this is clear. In some sense, allophonic variation does not introduce a faithfulness violation. Remember that McCarthy (2003b) tried to establish the same point for syllabification: this is never distinctive, therefore Cumulativity does not depend on it. In the case of allophonic vari-
ation, the variation is not distinctive either. But since ‘distinctivity’ in itself is not a primitive notion in OT (it is derivative of the fact that all relevant markedness constraints dominate all relevant faithfulness constraints), it is not easy to see how we could establish this restriction within Sympathy Theory – although obviously the fact that I do not see how it can be done does not mean that it is impossible to do it. Within LP phonology, we could try to restrict the relation between the constraint rankings of the lexical and the postlexical phonology. Itô & Mester (2002) suggest that one restriction might be that if a contextfree markedness constraint is below faithfulness in the postlexicon (i.e. some segment type is ruled out phonetically), it will not be ranked higher in the postlexicon (i.e. the postlexicon will not disallow segment types that are allowed lexically). Notice that such an assumption would

We now return to the Japanese case. Itô & Mester (2002) argue that the constraint set in (??) is mistaken (cf. footnote ??), for instance since it has to presuppose that in the world of context-free markedness, we find \(*g\gg\eta\), while in the world of context-sensitive markedness we find \(*[\text{PW}g]\gg*[\text{PW}\eta]\). They also argue that the ranking \(*\eta\gg*g\) is more plausible from a universalist perspective. For this reason they assume a slightly different constraint ranking to understand the basic alternation, involving a different context-sensitive constraint:

\[(41)\]
\begin{enumerate}
  \item \textit{Post-lexical ranking:} \textit{\(*VgV\gg\eta\gg*g\gg\text{IDENT-[nasal]}\)}
\end{enumerate}

This again gives us allophony, since the relevant markedness constraints outrank faithfulness. Itô & Mester (2002) assume that this is the ranking for the postlexical component. You can check for yourself that it gives the right results in the morphologically simplex cases (\textit{geta}, \textit{kanji}). In order to explain the opaque behaviour in Rendaku, Itô & Mester (2002) assume that the ranking in the lexicon is as in (42a) (the reranking conforms the restriction we imposed above). We then get a tableau for the Lexical component as in (42b) and (42c):

\[(42)\]
\begin{enumerate}
  \item \textit{Lexical ranking:} \textit{\(*\eta\gg*VgV\gg*g\gg\text{IDENT-[nasal]}\)}
  \item \textit{OCP\gg\text{RM}\gg\text{IDENT-[voice]}\)}
\end{enumerate}

<table>
<thead>
<tr>
<th>Form</th>
<th>\eta</th>
<th>OCP</th>
<th>RM</th>
<th>VgV</th>
<th>\eta</th>
<th>IDENT-[nas]</th>
<th>IDENT-[vc]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/saka-tope/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/saka-(tope)]</td>
<td>(\eta)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[saka-doge]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/saka-(tope)]</td>
<td>(\eta)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
c.

The output of the lexical component thus is \textit{saka-toge} regardless of the input specification. If we input this to the postlexical grammar (which thus is not subject to Richness of the Base), we get the following:

\begin{tabular}{|c|c|c|c|c|c|}
\hline
/saka-toge/ & \textit{*VgV} & \textit{g} & \textit{g} & \textit{IDENT-[nas]} & \textit{IDENT-[vc]} \\
\hline
\textit{saka-toge} & \textit{?} & \textit{*} & \textit{?} & \textit{?} & \textit{?} \\
\hline
\textit{saka-doNe} & \textit{?} & \textit{?} & \textit{*} & \textit{?} & \textit{?} \\
\hline
\end{tabular}

What about the candidate \textit{[saka-doNe]}? \textit{Itô & Mester} (2002) claim that the fact that it does not occur, shows that we need a further reranking: \textsc{Realize-Morpheme} now should be ranked below \textit{IDENT-[voice]}, expressing the intuition that Rendaku is no longer operative (since Rendaku is a morpheme, and its realisation is no longer required).

There is also an alternative, which \textit{Itô & Mester} (2002) do not discuss but which does not seem to be implausible from the point of view of Lexical Phonology: we could assume that the Rendaku morpheme disappears altogether. It may have been the input (and output) of the lexical component, but that does not mean it is still in the input of the postlexical component. In the first place, it is an established result that postlexical processes are insensitive to morphological structure. One way of establishing this is by assuming this structure is no longer visible to it, e.g. because it has been erased from the input. Secondly, if the Rendaku morpheme indeed consists of a feature \textit{[+voice]}, as \textit{Itô & Mester} (2002) suggest, this feature is floating in the output representation of the lexicon; it is possible that Stray Erasure applies between the lexicon and the postlexicon. In other words, we do not have to assume that the output of the lexicon and the input of the postlexicon are identical, just as we do not have to assume that the phonetic representation is exactly identical to the output of the (postlexical) phonology. Various automatic cleanup rules could have applied.

Actually, both within Correspondence Theory and within Containment Theory, we need to assume that there is some cleaning up between the two components. There is no reason, for instance, why the whole original input of the lexical component should be available to the postlexicon, even though both theories assume in one way or another that this input is still available in the phonological output. It could be then, that the the transfer of lexicon to postlexicon has further properties, such as deleting all the morphological structure. The postlexicon then becomes closer to the phonetic structure, which is a good thing. Obviously, any property of this transfer must be auto-
matic and not language-specific, because we do not want to keep to the principle of OT that all the explanation of differences between languages should come from differences in constraint ranking.

Itô & Mester (2002) also briefly discuss the ‘mimetic’ reduplication in Japanese (cf. (33)), where we do not have Rendaku. Another property of this is also that the initial ɣ in e.g. garu-garu does not turn into η. On p. 25, we have seen how this can be analysed as an instance of ‘underapplication’ of the process turning ɣ to η intervocally. The analysis there crucially relied on the fact that we have a constraint against word-initial η, however; a constraint which we have just abandoned. Yet Itô & Mester (2002) point out that a different analysis is also available, viz. one in which the two parts of the mimetic compound form an independent phonological word, and the constraint *VɣV only applies within the domain of such a word. The constraint then is inapplicable to either part of the compound, so that [garu-garu] is optimal, which is the required result. (In non-mimetic compounding, there is evidence that the two parts of the compound are not independent words, and here we do find normal application: kuni-Nuni ‘various countries’.

We now turn to another language, we have discussed above, Palestinian Arabic (section ??). As we have seen above, Kiparsky (2000) has pointed out a number of problems regarding such an analysis in terms of OO Faithfulness. The most important problem is that it is hard, if not impossible to give a satisfactory definition of the notion ‘base’ within such a theory, because it does not provide us with a good view of the interaction between phonology and morphology. In contrast, LP is build to give such a view, so we might expect it to fare better with respect to these facts.

Kiparsky (2000) notes, for instance, that an OO Faithfulness account does not relate the opacity of stress to anything else. In particular, it misses the generalization that all processes of word phonology ignore epenthetic vowels. For example, closed syllables are shortened even though postlexical epenthesis opens them,

<table>
<thead>
<tr>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. /jâaf-at/ [jâaf-at] ‘she saw’ (transparent retention of length)</td>
<td>a. /rubat-at/ [rubat-at] ‘she fastened’ (spreading blocked by a)</td>
</tr>
<tr>
<td>b. /jâaf-t/ [fîfît] (*[jâafît]) ‘I saw’ (opaque shortening)</td>
<td>b. /rubat-t/ [rubatît] ‘I fastened’ (opaque spread across i)</td>
</tr>
</tbody>
</table>

and epenthetic i differs from underlying vowels in not blocking the spread of “emphasis” (Ervin (1963, p. 83), for Iraqi Arabic).

Kiparsky (2000) mentions that Sympathy Theory suffers the same problem: it may be able to describe the facts, but it does not explain why epenthesis is invisible to all of word phonology. He therefore proposes to differentiate two levels within the lexical component, a stem level and a word level. This may seem like an arbitrary move, but Kiparsky (2000) claims that
The categories “stem” and “word” are special in being anchored in the universal prosodic hierarchy, their status in UG is comparable to the status of such categories as “noun” and “verb”. Unlike classical rule ordering theory and sympathy theory, where the orderings and rankings respectively are unconstrained, this imposes a small upper limit on the depth of opaque interactions. The difference between subject endings and object clitics is that the former are attached at the stem level and the later at the word level. We thus have the following derivations:

(44)  stem level  fıhim  fıhim  fıhım-na  
       word level  fıhim  fıhim-na  fıhım-na  
       he understood  he understood us  we understood

The idea thus is that stress is assigned at the stem level, and does not change on the word level; on the former, faithfulness of stress constraints have to outrank prosodic well-formedness.

All in all, the points in favour of an LP version of OT seem quite strong: it provides some level of opacity which is not extremely strong, but still available. This does not mean that there are no problems. We have already noted that the differences in ranking between components need to be restricted in a principled way — we do not want to predict grammars which combine the lexical phonology of English with the postlexical phonology of Chinese. Second (and maybe connected to this), we have also seen that we probably want to still restrict the power of the model, e.g. to exclude cases of DY derivation.

There are also a few other problems with LP, which seem to point in a different direction, viz. that the theory may be too restrictive. For instance, Anttila et al. (2004) discuss some facts from Singapore English. In this dialect, various process are active which serve to avoid consonant clusters, including Epenthesis (45a) and Metathesis (45b). Metathesis opacifies Epenthesis, however: the form in (45c) has a geminate [s:] rather than an epenthetic vowel.

(45)  a. Epenthesis  lap/s+z/  →  lap[saz]  
       b. Metathesis  li/sp/  →  li[ps]  
       c. Epenthesis made opaque  li/sp+s/  →  li[ps] (*li[ps])

Within LP, we would assume that Metathesis comes at a later level than Epenthesis. Anttila et al. (2004) note that Epenthesis must be a word-level (or later) process for morphological reasons: it is triggered by word-level suffixes, e.g. /laeps+z/  →  [laepsaz]. This implies that Metathesis must be postlexical. Yet Anttila et al. (2004) show that there are at least three reasons to reject this assumption. First, Metathesis is sensitive to whether the following segment is a consonant or a vowel when it occurs in the middle of a word, but
1.11. Coloured Containment

it is not sensitive to this distinction when it occurs at the end of a word (followed by another word). Secondly, there is another process in Singapore English (Fricativization, \( p \rightarrow f \)) which counterbleeds Metathesis (\( lisp \rightarrow lips \rightarrow lifs \)). Fricativization would thus have to be ordered after Metathesis, but nothing comes after the postlexical level. And finally, metathesis does not seem to occur postlexically in any other known language of the world (which in itself is an interesting observation).

It should be noted that there is no alternative which really works well. OO Faithfulness has the problem that the crucial property is in the input, not in some output (since both \([lips]\) and \([læps]\) exist). Anttila et al. (2004) offer a solution in terms of Comparative Markedness: Epenthesis resolves underlying sibilant clusters (‘old’ clusters), but not ones created through Metathesis (‘new’ clusters).

\[
\begin{align*}
\text{(46) a. } & \text{lap}/s+z/ \rightarrow \text{lap}[s@z] \text{ (s-z is ‘old’, repaired by Epenthesis)} \\
\text{b. } & \text{li}/sp+z/ \rightarrow \text{li}[psz] \rightarrow \text{li}[ps:] \text{ (s-z is ‘new’, not repaired)}
\end{align*}
\]

This works, but only on a crucial assumption: that the order of affixes is also present underlyingly. This is a questionable assumption, however. For instance, a standard analysis of OT has it that the order of affixes can be determined by ranked constraints.

1.11 Coloured Containment

It is now time to evaluate what we have seen so far. Broadly speaking, there are two theories of faithfulness: one which is based on the assumption that the input is part of the output, and another is based on the idea that input and output are completely independent representations. The first theory is instantiated by Containment Theory, and it has many problems; the most difficult puzzle is that does not offer a satisfactory theory of epenthesis or of the insertion of features. On the other hand, the theory is constrained by principles such as Containment (definition 7.2 on p. 115) and Consistency of Exponence (definition ?? on p. ??).

The second theory, instantiated by Correspondence Theory, seems much less constrained: any input can correspond to any output. We have concluded in chapter ?? that such a move can only be justified if there is independent evidence for correspondence relations in natural language. The last ten years have seen the search for such independent evidence. But we have also seen that all of the cases of Correspondence Theory which have been put forward, have also been questioned on serious grounds:

1. OO Faithfulness is fraught with problems, mostly because it is not clear how it can be constrained. Furthermore, most analyses can be replaced
1.11. Coloured Containment

by IO Faithfulness constraints under the assumption of Lexical Phonology. Such an analysis seems more constrained, and better adapted to well-established theories of morphology. For LP, we do not need Correspondence Theory, however.

2. Similarly, Sympathy Theory can be replaced at least in part by theoretical instruments such as Comparative Markedness, which does not necessarily depend on Correspondence Theory, and (again) on Lexical Phonology. Furthermore, we will see below (section 1.13) that within Containment Theory there also are other devices for handling apparent cases of opacity.

3. The best candidate for a correspondence relation outside input-output faithfulness therefore still is found in reduplication. But we have seen there that alternative analyses are available in this case. Let us for instance, return briefly to the case of Javanese overapplication in (30) repeated here for convenience.

(30) a. putul ‘cut (n.)’
   b. pang-putul ‘that used for cutting’>[pamutul]
   c. pa-mu-mutul ‘a cutting in quantity’ (reduplication; *pa-mu-putul)

Remember that the prefix pang nasalises the immediately following obstruent; but in reduplicated forms it nasalises the first segments of both the base and the reduplicant.

In the preceding sections, we have developed an alternative terminology for describing these forms: we could say that the nasalisation has applied both opaquely (to the base mutul) and transparently (to the reduplicative prefix mu). We have seen that Inkelas & Zoll (2005) propose that the reduplicant is not a prefix but an infix, and faithfulness applies in two steps (example (36), also repeated here):

(36) putul ↓ + pang
    ↓ + RED
    pang-mu-mutul

We have seen now that there is faithfulness machinery independent of Correspondence Theory that is also able to handle this, for instance Lexical Phonology. (Inkelas & Zoll 2005) use a theory of cyclic representations as discussed in section ??.) It would be worth finding out whether all cases of overapplication could be subsumed to such an analysis.

We have seen that a case of underapplication could potentially also be handled in a different way: nasalisation of velar stops does not apply to
1.12 Morphological colouring

In any case, we have seen that there is not very strong evidence in favour of Correspondence Theory. My conclusion from this is that we need a more restricted theory. We have seen that Containment Theory has many problems, at least in its PARSE/FILL implementation.

Let us try to find out where these problems lie. We have already established that they basically have to do with the theory of epenthesis. In the implementation of Prince & Smolensky (1993), this is affected by FILL constraints: epenthetic segments are empty and this is why they are disallowed. There is no real problem with the PARSE component.

In section ?? above, we have seen that a potential theory of faithfulness based on Comparative Markedness has the inverse problem. It is very easy to disallow epenthesis (by a ‘new’ version of the constraint against all structure, *MSTRUC, which thus disallows all inserted structure), but it is not clear how we can ban deletion in these terms — we would need to count those violations in the input structure which are not present in the output, but this is a mechanism that does not seem necessary for anything beyond faithfulness.

We could reach a solution if it were possible to combine the two approaches. At the end of section ??, we have already noted that Containment Theory would be able to incorporate Comparative Markedness if it would be possible to distinguish old from new structure. My claim is that this is indeed possible, without even extending the general framework already proposed in Prince & Smolensky (1993), in particular, by taking the principle Consistency of Exponence (def. ?? on p. ??) seriously. What does it mean to say that ‘No changes in the exponence of a phonologically-specified morpheme are per-
mitted’? This can only have a meaning if the ‘exponence’ of morphemes are visible in the surface structure.

One way to visualize this, is to assume that every morpheme has a specific colour, that every element in the underlying specification of morphemes also has this colour, and that Gen cannot change the colour of segments. Since colours are a little bit hard to use in modern print (and we would need a large number of colours for natural languages with large numbers of morphemes) we will use the descriptive notation of subscripts. (In return, we no longer need the subscripts for correspondence relations, obviously.)

Epenthetic material will now be ‘colourless’: it doesn’t belong to any morpheme and therefore it does not have a subscript (or a subscript $\emptyset$). Constraints against insertion will therefore be constraints against colourlessness. Obviously, such constraints can be combined with a containment model. Output structures in this ‘Coloured Containment’ (CC) model will look as follows (cf. (4) on p. 6 and (??) on p. ??):

\[
\begin{array}{c}
\sigma \\
\begin{array}{c}
t_a \\
\alpha \\
k_a \\
p_a \\
\emptyset
\end{array}
\end{array}
\]

(47)

As in ordinary Containment, we know that $k_a$ is deleted because it has not been parsed in the phonological structure. Like in Correspondence Theory, the epenthetic vowel is phonologically fully specified (with whatever the features of schwa are); but like in Containment Theory, it is also empty, viz. in its morphological affiliation. The most important faithfulness constraints are thus:

\[
(48) \quad \begin{align*}
a. \ & \text{PARSE-} \phi(\alpha): \text{The morphological element } \alpha \text{ must be incorporated} \\
& \text{into the phonological structure. (No deletion.)} \\
\text{b.} \ & \text{PARSE-} \mu(\alpha): \text{The phonological element } \alpha \text{ must be incorporated} \\
& \text{into the morphological structure. (No insertion.)}
\end{align*}
\]

We have to assume that the phonetic component (and possibly cleanup rules between the lexical and the postlexical component) will apply Stray Erasure, and erase the morphological colours, so that the phonetic output will look (approximately) as:

\[
\checkmark \ t \ a . \ p \ \emptyset
\]

Notice that we did not distinguish here between segments and (autosegmental) features. It is not necessary to do so in this model, although it would be possible (we could restrict the definitions above to phonological segments and then introduce the equivalent of IDENT constraints). It would also be possible to model some of the more specialized constraints, such as ANCHOR
I-ANCHORING-(L/R) in CC
No phonologically unparsed elements may occur at the left/right to the input string.

O-ANCHORING-(L/R) in CC
No morphologically unparsed elements may occur at the left/right to the input string.

These constraints have the flavour of positional markedness to them, and in general, this is how they can be treated. CC furthermore can also incorporate most of the findings of Comparative Markedness, as becomes clear if we take another look at the example from the Makassarese example (??), repeated here:

(??) a. /rantas/ [rântasa?] ‘dirty’
     /jamal/ [jâmalâ?] ‘naughty’

b. /lombo/ [lîombo] ‘big’

In order to distinguish the a- and b-cases, McCarthy (2003a) proposes to distinguish two cases of the constraint FINAL-C. In terms of CC, we could do this in the following way:

✓ _N_ FINAL-C: No morphologically unparsed vowel at the end of the word.
✓ _O_ FINAL-C: No morphologically parsed vowel at the end of the word (i.e. words should end in a syllable boundary).

In other words, _N_ FINAL-C is a special case of O-ANCHORING-(R) in this approach; _O_ FINAL-C is a case of what used to be called ALIGNMENT (McCarthy & Prince, 1995b).

It might be instructive to look more precisely at the CC implementation of CM. We will return to another case of derived environment effects, Vowel Harmony in Turkish. Remember from (??) that this language has vowel harmony progresses for the features [±back] and [±round]. We have not mentioned yet that many stems are disharmonic:

     _muzip_ ‘mischievous’, _buket_ ‘bouquet’, _mebus_ ‘MP’

Most of these stems are loanwords. We may observe that loanwords are often adapted to other aspects of Turkish phonology; for instance, in pairs like _biftek-biftei_ ‘steak - his steak’, /k/ disappears in intervocalic position, in accordance with a general rule, and in disharmonic _kitap_ ‘book’ (<Ar. kitab), the /k/ gets fronted because of the vowel and the final plosive is devoiced. The question is why this did not happen to root-internal vowel harmony, and it seems sensible to say that vowel harmony is _blocked_ in these cases.
However, colloquial varieties of Turkish show epenthesis: if a word (again, a loanword) starts with a cluster of consonants, an epenthetic vowel is inserted. It is important to see that this vowel is harmonic:

(50) | careful form | colloquial form |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>‘fetters’</td>
<td>pranga</td>
</tr>
<tr>
<td>‘prince’</td>
<td>prens</td>
</tr>
<tr>
<td>‘test’</td>
<td>prova</td>
</tr>
<tr>
<td>‘announcer’</td>
<td>spiker</td>
</tr>
<tr>
<td>‘cruiser’</td>
<td>kruvazör</td>
</tr>
</tbody>
</table>

There is a relation between the existence of disharmonic stems and the possibility that affixes and epenthetic vowels can receive harmony: within a stem a feature cannot be given to ‘old’ segments. We can only make a vowel harmonic if it is in some sense a ‘new’ vowel.

(51) a. pilot, otel, kruvazör: none of the vowels is new, they have always belonged to the same word, therefore they need not be harmonized
b. pilot+lar, otel+ler: the vowel in the suffix is new, it should be harmonized
c. kruvazör: the epenthetic vowel is new, it should be harmonized

The conclusion of this is that new vowels should be harmonized. Old vowels do not need to be, although over the course of the years they will tend to do so. (Intuitively, this is why disharmonic roots are loanwords: they did not have enough time yet to adapt themselves to the phonological system of Turkish.)

A problem with the CM framework in cases like this is that it is completely symmetric. We predict that there will be cases of ‘anticyclicity’. Thus, assume a hypothetical language L with roundness harmony, a stem kitap, and suffix -ler/lar and -en/an. Given high-ranking \( O_{\text{SPREAD}} \) (and low-ranking other \( \text{SPREAD} \) constraints), we would have a language Turkish \( \neg \) with the following properties:

- Underived form: kitep (\( O_{\text{SPREAD}} \) is applicable, since \( N_{\text{SPREAD}} \) is not).
- Derived form: kitel-lar (\( O_{\text{SPREAD}} \) applies to stem element, but not to suffixes).
- Derived form: kitel-lar-an (\( O_{\text{SPREAD}} \) applies to kitepler, but not to the whole word)

I am not aware of any language that has this type of harmony, and, intuitively, it seems quite absurd. \( O_{\text{SPREAD}} \) has more of these strange ‘anti-cyclic’ consequences, where things happen only to forms once they are fully embedded in other forms. For instance, a language which allows voiced obstruents
in the middle of an underived word, but not if an affix is attached to it, giving alternations like pobin/popini; or making all syllables open in a derived environment, giving alternations like pambin/pabina, etc.

I therefore propose another option: general markedness constraints such as \textsc{Spread} do not come in two versions: they do not see the difference between old and new structure. This also seems more plausible. Why would constraints on surface well-formedness care about the origin of the material it has to evaluate? Rather, there is a range of constraints which regulate the relation between morphological structure and phonological structure. The basic idea behind all of these constraints is the following (the constraint is named after Kaye (1974)):

\begin{definition}[Morphological recoverability (MR)]
Phonological structure mirrors morphological structure as closely as possible.
\end{definition}

In the case at hand, one such MR constraint could be the following:

\textbf{\textsc{Integrity}} (first definition): Old vowels may not contain new material.

We find evidence for this in Icelandic. In this language, /a/ turns to [ò] when it is followed by an [ii]. The epenthetic vowel happens to be [ii] as well, but this vowel has no effect on the preceding /a/. Notice that in this case, the epenthetic vowel is not a potential undergoer of the spreading, but a potential trigger.

\begin{enumerate}
\item a. \textit{barn} ‘child-ACC.SG.’ \quad \textit{börn-um} ‘child-DAT.PL’
\textit{dag} ‘day-ACC.SG.’ \quad \textit{dō-um} ‘day-DAT.PL’
\textit{vatn} ‘water-ACC.SG.’ \quad \textit{vōtn-um} ‘water-DAT.PL’
\textit{tala} ‘I speak’ \quad \textit{tōl-um} ‘we speak’
\textit{baka} ‘I bake’ \quad \textit{bōk-um} ‘we bake’
\textit{kaldla} ‘cold-ACC.SG.’ \quad \textit{kōld-um} ‘cold-DAT.PL’
\item b. \textit{dag-ur} ‘day-NOM.SG.’ (/dag/+/r/)
\textit{snarp-ur} ‘rough-MASC.NOM.SG.’ (/snarp/+/r/)
\end{enumerate}

Notice that we have to differentiate in this case between vowels which are new because they are in a suffix and vowels which are new because they are epenthetic; we will not go into this difference here. The most important thing is that a language Icelandic\textsuperscript{\textasciitilde} with the following properties does not exist:

\begin{itemize}
\item Underlying vowels do not act as triggers for spreading (/barn/ + /um/ >[barnum]),
\item but epenthetic vowels do (/agr/ >[ögur]).
\end{itemize}

The non-existence of Icelandic\textsuperscript{\textasciitilde} can be explained in the same way as the non-existence of Turkish\textsuperscript{\textasciitilde}: the inverse of the constraint \textsc{Integrity} (which would read ‘New vowels cannot contain old material’) is not part of the universal inventory of constraints. Of course, this begs the question: why does
the principle of Morphological Recoverability lead to INTEGRITY and not to \( \neg \)INTEGRITY? We will briefly return to this in section ??.

### 1.13 Opacity and output representations

We have seen that most of the arguments for the power of Correspondence Theory derive from opacity, especially if we consider that many cases of over-application and underapplication in reduplication can also be derived from it. In this section, we will show that CC is also quite powerful and can handle quite a large number of phenomena, albeit not in a uniform fashion. Generally speaking, we have (i) INTEGRITY and other Recoverability constraints and (ii) enriched representations at our disposal.

In the first place, we have already seen that CC can mimick many aspects of CM theory. This applies to opacity effects as well. McCarthy (2003a) shows that CM can deal with cases of counter-feeding opacity such as the following, from Barrow Inupiaq (an Inuit language):

\[(53)\]

a. Palatalisation after /i/

<table>
<thead>
<tr>
<th>Stem</th>
<th>-lla</th>
<th>‘be able’</th>
</tr>
</thead>
<tbody>
<tr>
<td>/niiri/</td>
<td>[niiriλla]</td>
<td>‘eat’</td>
</tr>
<tr>
<td>cf. /sisu/</td>
<td>sisulla</td>
<td>‘slide’</td>
</tr>
</tbody>
</table>

b. No palatalisation after /ı/

| /tiıjr/ | [tiıjr] | ‘take flight’ |

Palatalisation of /l/ to [λ] is opaque in this language, since it does not apply after a ‘derived’ [i] (one which is the result of fronting an epenthetic i). Within CM, we can analyse this as the result of an ‘old’ version of a constraint against palatalisation: every [l] next to an [i] is a violation, if /li/ were adjacent already underlingly. In CC, we would have one Palatalisation constraint, but INTEGRITY would do the required work. It would state that old l’s may not acquire new palatalisation features.

Also, many apparent cases of opacity actually seem to disappear if we assume a somewhat more sophisticated view of phonological representations, taking into account the literature on prosodic organization, autosegmental structure and the interaction with morphological and syntactic boundaries. One representation, which is enriched by independently necessary elements, may then do the work of two poorer representations.

In the Antwerp dialect of Dutch, for instance, we have a process velarising a nasal consonant in coda position. We also have a process shortening a vowel before the resulting velar nasal (Taeldeman, 1982). Both processes can be seen at work in the examples in \[(54)\]. Interestingly, the velarisation process only applies to words with underlingly long vowels, and not by words which have short vowels already underlingly, as \[(54b)\] shows:
1.13. Opacity and output representations

(54) a. i. grune ‘green’ \[\text{gryn}\] \(\sim\) gruun ‘green’ \[\text{gryn}\]
    ii. schoenen ‘shoes’ \[\text{sxn}\] \(\sim\) schoen ‘shoe’ \[\text{sxn}\]

b. i. kin ‘chin’ \[\text{kin}\] \(\sim\) tien ‘ten’ \[\text{ti}\]
    ii. zon ‘sun’ \[\text{zon}\] \(\sim\) zoon ‘son’ \[\text{zon}\]

This process clearly looks opaque: we could claim for instance that velarisation only applies after long vowels, but this vowel length is not visible on the surface. On closer inspection, there is evidence that the velar nasal, different from the other nasal consonants, is a moraic segment in varieties of Dutch (Trommelen [1982]; van Oostendorp [2001]; van der Torre [2003]). This could explain, for one thing, the fact that velar nasals can only occur after short vowels: we could posit a bimoraic maximum on syllables.\(^5\) A velar nasal after a long vowel would then be prohibited:

\[
\begin{array}{cccc}
\text{bang ‘afraid’} & \text{ban ‘ban’} & \text{baan ‘ban’} & \text{*baang} \\
\mu & \mu & \mu & \mu & \mu & \mu & \mu & \mu & \mu
\end{array}
\]

(55)

b a n b a n b a n b a n

There undoubtedly are other ways to capture the same intuition: that both long vowels and velar nasals need space in the syllable and that if we would have both, this would be too much. Under the one chosen here, the opacity of the Antwerp Dutch velarisation process disappears, if we assume that there is a strong faithfulness requirement on the number of mora’s in this dialect — or in rule-based terms, that we are not allowed to insert any mora’s in the underlying representation. A change from /\text{ti:n}/ — two underlying mora’s attached to the vowel — to [\text{ti}n] (two surface mora’s, one for the vowel and one for the nasal) is then allowed, but a change from /\text{ki:n}/ (one underlying mora attached to the vowel) to [\text{ki}n] (two surface mora’s) is not. If we allow ourselves to introduce a few ad hoc constraints to make things work technically, an OT analysis might then run along the following lines. We need the constraints in (100), and the ranking in (101) in order to get the tableau in (102):

(56) a. i. FAITH(\(\mu\)): Do not add or delete mora’s
    ii. VELAR: Nasal consonants in coda position shoul be \(\eta\).
    iii. *\(\mu\mu\mu\): No trimoraic syllables

\(^4\)Apart from some cases of optional schwa deletion, the form without schwa can only be used in the singular neuter of adjectives in indefinite noun phrases. The form with schwa can be used in all other inflections (plural or non-neuter, or both).

\(^5\)It is sometimes assumed that the reason for this restriction is that velar nasals underlyingly are /\text{ng}/ or /\text{nG}/, just like this is assumed for English. The Antwerp facts actually show that this analysis cannot be correct: here we find the same restriction but there is absolutely no possibility for postulating an underlying obstruent.
b. Faith(µ) ≫ Velar

c. i. 

<table>
<thead>
<tr>
<th>/tin/</th>
<th>Faith(µ)</th>
<th>*µµµ</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>[tin]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[t:in]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[t:i:n]</td>
<td>*!</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>[tin]</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

ii. 

<table>
<thead>
<tr>
<th>/kin/</th>
<th>Faith(µ)</th>
<th>*µµµ</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>[kin]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[k:in]</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>[k:i:n]</td>
<td><em>!</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ki:n]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All three constraints are known from the literature. No special mechanisms, such as rule ordering, Sympathy Theory, levels, etc., thus are necessary for Antwerp Dutch, given the appropriate assumptions on the representation of the velar nasal.

An alternative solution to some other cases of opacity, which uses even richer representational devices has been proposed under the rubric of ‘turbidity’ of Goldrick (2000). His analysis is based on the opaque vowel length we find in Luganda and which has been discussed within Cumulativity Theory on p. ??.

The relevant facts are in (??), repeated here:

(??)  /ku+linda/ → [kuli:`da] ‘to wait’
/ku+ntu/ → [ku:`tu] ‘person’
/ba+ntu/ → [ba:`tu] ‘people’

Goldrick (2000) proposes to picture the output phonological structure of muntu in the following way:

\[
\begin{array}{c}
\mu
\\ \mu
\\ \mu
\\ \mathbf{\mu}
\\ \mathbf{\mu}
\\ m
\\ u
\\ n
\\ n^t
\\ u
\end{array}
\]

The [n] and [u] maintain different relations with the mora. The relation between n and mora is called Projection, and Goldrick (2000) notes that this is roughly equivalent to what is called licensing. It is denoted by an arrow pointing upwards. The relation between the vowel and this mora is called Pronunciation; it is denoted by an arrow pointing downwards. In the default case, Projection and Pronunciation relation will be maintained by the same pair of elements (this is true for the other mora’s in this representation; the
Across morphological boundaries

2 Across morphological boundaries

2.1 Asymmetries between prefixes and suffixes

Asymmetries in phonological behaviour between types of affixes are not uncommon in languages of the world. For instance, prefixes in a given language may behave quite differently from suffixes. Particular classes of suffixes (or prefixes) may show different types of behaviour as well. For many Germanic languages, it has been claimed that we should distinguish between so-called Class I and Class II suffixes (or between morpheme boundaries + and #, or between lexical levels I and II, or between ‘cohering’ and ‘non-cohering’ suffixes). In many cases, the two classes of suffix have completely different shapes. For instance, Class I suffixes are typically vowel-initial and at most monosyllabic, whereas Class II suffixes often are consonant-initial and have more material than fits in one syllable. The issue arises whether we should set up morpheme structure constraints to account for these differences, or we should rather derive the morphological status from the phonological form.
The goal of this class is to show that morphological diacritics are mostly unnecessary in the phonology of Dutch. Differences in phonological behaviour of different morphemes can be derived from the underlying phonological shape of these morphemes, provided we have a theory of violable constraint interaction such as Optimality Theory.

Before laying out the theoretical apparatus in full, we will first turn to one set of examples illustrating the topic of interest: syllabification across morpheme boundaries. Tautomorphemic sequences of a consonant and a vowel in Dutch are syllabified together (58a), as might be universally the case (cf. Piggott (2002) for recent discussion). The same happens if the consonant is at the end of a stem and the vowel is initial in the following suffix (58b). However, the picture changes if the consonant belongs to a prefix and the following vowel to another prefix or to a stem. In this case, the syllable boundary will fall between the consonant and the vowel (58c).

(58)  
   a. *ode ‘ode’ [o.da]
   b. *er+en ‘to honour (+INF) [e.ran]
   c. *ont+eer ‘dis+honour’ [ant.er]

It should be noted that the syllable boundaries assigned here do not just correspond to native speakers’ judgments, but they also have a clear effect on phonological alternations that are dependent on syllabification. The most important one of these is a schwa-zero alternation found in Dutch, and exemplified in (59) below:

(59)  
   a. elite elite+air
       [e.li.ta] [e.li.ta] / *[e.li.ta.rr]
       ‘elite’ ‘snobbish’
   b. adem be+adem
       [a.dam] [ba.a.dam] / *[ba.dam]
       ‘breathe’ ‘breathe upon’

In Dutch monomorphemic forms we never find a schwa immediately preceding another vowel. This restriction can be understood as a result of syllable optimization: schwa as a vowel has minimal feature content so that we may assume that it can be deleted relatively easily; it does not have a lot of underlying features that surface structure should faithfully reflect. In particular, we may assume that the faithfulness requirements demanding schwa to surface (?) are ranked below the constraint ONSET. This can be observed in the derivation of the following (hypothetical) underlying form /mɔan/:
2.1. Asymmetries between prefixes and suffixes

In affixed forms we find again an asymmetry between prefixes and suffixes, as can be seen in (59). This asymmetry can be reduced to the syllable structure asymmetry we have just analyzed. The schwa cannot be deleted if it ends a prefix, because the resulting surface syllable would cross a prefix-stem boundary, but the schwa at the end of a suffix can be deleted under the appropriate circumstances, because syllabification over a stem-suffix boundary is not blocked.

It is easy to think of a number of unsatisfactory solutions to this problem:

- **Suffixes are morphologically attached before prefixes are.** The problem with this solution is that it is arbitrary; furthermore in a case like ont+eer+en (dis+honour+InfL), the suffix is inflectional and the prefix is derivational, which makes it somewhat harder to adopt this solution.

- **Dutch adopts a ranking ALIGN-L ≫ ALIGN-R.** This solution is also arbitrary; does not take into account the fact that this ranking is universal, or otherwise it merely postulates this universality.

- **“Processing pays special attention to the left edge.”** Even this solution is to some extent arbitrary, since it does not explain why this is the case; processing also pays attention to the right edge.

We try to find a more principled explanation in terms of the concept *integrity*. We assume that every syllable has a head, which is a is the most sonorous segment of the syllable. Our second assumption is that instead of directional Alignment constraints on the morphological interface, we have the constraint $\text{PR} \approx \text{LX}$ (Prince & Smolensky, 1993) which demands that the edges of morphological constituents should coincide with those of prosodic constituents and vice versa, without stipulating a difference between left and right boundaries:

\[
\begin{array}{|c|c|c|}
\hline
/\text{maan}/ & \text{ONSET} & \text{FAITH-\textit{\textdagger}} \\
\text{ma.an} & *! & * \\
\text{\#maan} & & * \\
\hline
\end{array}
\]

In essence, $\text{PR} \approx \text{LX}$ is a symmetrical version of asymmetrical ALIGNLEFT and ALIGNRIGHT; like these constraints, it bans boundaries that do not cooccur. Unlike these, it does not distinguish between the left-hand side of the word and the right-hand side of the word. In the analysis of Dutch word-stress above, an asymmetric constraint ALIGNRIGHT has been used, but only in the context of aligning prosodic structure to other prosodic structure. Our claim is that even though purely phonological constraints on foot placement and the like can potentially refer to left and right edges of words, constraints on
the interface between phonology and morphology cannot. A reason for this may be that the notions ‘left’ and ‘right’ are relevant only for phonology, not so much for morphology or syntax, in which other notions, such as hierarchy and embedding, play a role.

A last assumption is that phonological segments have a morphological domain. Typically, this is the smallest morphological word to which they belong. We will demonstrate this on the example onteren ‘to dishonour’ in Dutch. This word consists of a nominal stem eer, a verbalizing prefix ont- and an inflectional suffix -en:

(62) \[ \left[ V \left[ V \text{ont} \left[ N \text{eer} \right] \right] V \text{en} \right] \]

The square brackets in this example indicate the boundaries of words: eer can act as an independent word, and so can onteer and onteren.6 Let us now consider the phonological domains of each of the three vowels. The /e/ in the stem eer has this stem as its morphological domain, since this is the smallest potentially independent word in which it appears. The vowel /o/ in the prefix has the derived form onteer as its domain, since the prefix is not a word in its own right. Finally, the schwa in the suffix has the whole word onteren as its domain, since this is the smallest independent word in which it occurs.

Slightly more formally we can now define the notion morphological domain in the following way:

(63) The morphological domain of a segment S is the smallest morphological word in which S occurs.

Next, we can define the morphological domain of syllables. Since syllables are headed, we can do this in terms of segment domains:

(64) The morphological domain of a syllable T is the morphological domain of the segment heading T.

Thus, in the example above, the domain of the syllable headed by /e/ is eer, the domain of the syllable headed by /o/ is onteer, since ont- is not a separate word, the domain of the syllable headed by schwa is onteren. With this theoretical apparatus set up, we can now propose a formalization for a constraint accounting for the difference between prefixes and suffixes:

(65) Morphological syllable integrity (INTEGRITY):
    All segments in a syllable should be in the same domain as that syllable.

---

6The fact that eer is written with only one \(<e>\) in the latter form, is a caprice of Dutch orthography.
∀ segment S: ∀ syllable T dominating S: the morphological domain of S ⊆ the morphological domain of T.

This constraint says, roughly, that all segments within a syllable should be in the same (smallest) word as the head of that syllable. To see how this works, consider once again our example onteren. The domain of the second vowel /e/ is the root, therefore all the segments in the syllable headed by this vowel should be in the root eer. The /t/ of the prefix is outside of this domain, therefore the syllabification *on-teren is not allowed. The domain of the schwa vowel in the suffix, on the other hand, is the whole word. The /r/ at the end of the root obviously is within this domain and therefore the syllabification onte-ren is allowed by INTEGRITY. In constraint tableaux (irrelevant morpheme boundaries have been omitted):

\[
\begin{array}{|c|c|c|}
\hline
\text{un+one+ish 'disputing'} & \text{INTEGRITY} & \text{ONSET} \\
\hline
\text{en} & * & * \\
\text{on} & ** & * \\
\hline
\end{array}
\]

Alignment is irrelevant in the cases at hand, and we may therefore assume that PR≈LX is ranked below the ONSET constraint. In all, we have three constraints, which may be ranked in the following three distinguishable ways:

\[
\begin{array}{|c|c|}
\hline
\text{INTEGRITY} & \text{ONSET} \\
\hline
\text{PR≈LX} & * \\
\hline
\end{array}
\]

2.2 Lexical levels

Following the original proposals of Booij (1977), most phonologists have assumed that there are two types of Dutch suffixes: ‘Class I’ suffixes and ‘Class
II’ suffixes in the original terminology. The following lists are copied from Booij (1977):

\[(68)\]

\[\begin{align*}
\text{a. Class I:} & \quad -aal /\text{al}/, -aan /\text{an}/, -age /\text{aze}/, -air /\text{a}:r/, -ast /\text{ast}/, -eel /\text{el}/, -eer /\text{:er}/, -eis /\text{es}/, -egge /\text{e}:g/, -ein /\text{ein}/, -erig /\text{erix}/, -erij /\text{erij}/, -es /\text{es}/, -esk /\text{esk}/, -eur /\text{eur}/, -eis /\text{es}/, -eian /\text{ijan}/, -ide /\text{id}/, -ief /\text{id}/, -ieik /\text{id}/, -iet /\text{id}/, -ieus /\text{id}/, -in /\text{in}/, -iseer /\text{id}/, -isme /\text{is}\text{ma}/, -ist /\text{id}/, -iteit /\text{id}/, -ei /\text{id}/, \\
\text{b. Class II:} & \quad -achtig /\text{a}:\text{xtix}/, -loos /\text{los}/, -ling /\text{lin}/, -baar /\text{b}\text{a}:r/, -dom /\text{d}\text{om}/, -heid /\text{h}\text{eit}/, -nis /\text{n}\text{is}/, -chap /\text{sx}\text{ap}/
\end{align*}\]

At least three phonological differences are supposed to be related to the distinction between Class I and Class II suffixation. One of these has to do with stress and we will return to this later; the other two criteria are the by now familiar syllabification and schwa deletion, which apply across the boundary between a stem and a Class I suffix, but not across a Class II boundary:

\[(69)\]

\[\begin{align*}
\text{a. Class I suffixes trigger resyllabification, Class II suffixes do not} & \quad (/\text{mohAmmEd} + /\text{an}/ \rightarrow /\text{mohAmmEdan}/, \text{vs.} /\text{halv} + /\text{lin}/ \rightarrow /\text{hAmlin}/) \\
\text{b. Class I suffixes trigger schwa deletion, but Class II suffixes do not} & \quad (/\text{sinodal} + /\text{al}/ \rightarrow /\text{sinodal}/, \text{vs.} /\text{warda} + /\text{los}/ \rightarrow /\text{warda}/)
\end{align*}\]

We can see that no resyllabification has applied in (69a) because the stem-final /v/ is devoiced: final devoicing does not normally apply to obstruents in the onset of a syllable.

Since prefixes trigger neither resyllabification nor schwa deletion, Booij (1981, 2002) assumes that all Dutch prefixes belong to Class II inherently. We have seen above that this somewhat arbitrary stipulation is no longer necessary if we adopt INTEGRITY.

Similarly, it seems that the distinction is superfluous for almost all of the suffixes as well. Since all Class I suffixes start with a vowel we get the desired result from the constraint ranking immediately, as we have in fact seen above:

\[(70)\]

<table>
<thead>
<tr>
<th>/mohAmmEd/ + /an/</th>
<th>INTEGRITY</th>
<th>ONSET</th>
</tr>
</thead>
<tbody>
<tr>
<td>mohammed+-an, ‘muslim’</td>
<td>![Mohammed+-an]</td>
<td>![Mohammed+-an]</td>
</tr>
</tbody>
</table>

For most consonant-initial suffixes, the miniature grammar developed until now also gives the correct output, but vacuously so, since both INTEGRITY and ONSET are irrelevant:
In this case, the cluster /ln/ is not even a potential syllable onset in Dutch so that we do not have to worry about the fact that we cannot distinguish between these two syllabifications in terms of alignment between morphological and phonological structure. If a suffix starts with a liquid (-loos, -ling), and the preceding stem ends in an obstruent, a potentially ambiguous situation arises. The word *werkloos* ‘idle’ (litt. ‘work-less’) could be syllabified either as *werk.loos* (respecting the boundary between stem and suffix) or as *wer.kloos* (satisfying the maximal onset condition).

The former option is actually chosen, and Booij (1977) takes this as evidence for his claim that consonant-initial suffixes belong to Class II: the syllabification rules of Class I do not apply in werkloos as they do in mo-hammedaan, therefore the two suffixes should be different. In the framework presented here, however, this line of reasoning does not hold. There is no single ‘process’ or ‘rule’ which syllabifies consonants into onset positions. Rather, there are several independent wellformedness constraints on the syllabified output structure. One such constraint is ONSET, but this is irrelevant in the case of werkloos, because it does not select between the two competing candidates. Therefore, another constraint becomes relevant, the Syllable Contact Law (SC):

\[(72)\] Syllable Contact (SC)
\[^{SC}\]C_i, C_j, where C_i is less sonorous than C_j.

This constraint is normally undominated in Dutch, giving syllabifications such as [ta.blo] rather than [tap.lo] for *tableau*. Yet it cannot be undominated in this case, because this would give us the incorrect results. We therefore need to find a constraint that can dominate SC. We have already seen this constraint above: \(PR\approx LX\):

\[(73)\]

<table>
<thead>
<tr>
<th>/vœyl/+/nis/</th>
<th>INTEGRITY</th>
<th>ONSET</th>
</tr>
</thead>
<tbody>
<tr>
<td>*vœ.lnIs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*vœ.lnIs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We should now establish the relative order between the two subrankings INTEGRITY ≫ ONSET and \(PR\approx LX\) ≫ SC. The order between ONSET and \(PR\approx LX\) is readily established: once we have another look at the vowel-initial suffixes we see that ONSET should dominate \(PR\approx LX\):
The rest of the ordering follows by transitivity. We have therefore established the following constraint ranking for Dutch:

\[
\text{INTEGRITY} \gg \text{ONSET} \gg \text{PR} \approx \text{LX} \gg \text{SC} \gg \text{FAITH-} \text{o}
\]

There is now one suffix left which is problematic for the account presented here. This is \(-achtig\) ‘-like’. The problem with this is that it seems to be the only vowel-initial suffix which does not belong to level I. It does not trigger resyllabification (76a) or schwa deletion (76b).

\[
\begin{align*}
\text{a. } & \text{roodachtig ‘reddish’ } \{\text{rot} \cdot \text{Ax}.t\text{Ix}, *\text{ro}.d\text{Ax}.t\text{Ix}\} \\
\text{b. } & \text{oranjeachtig ‘orange-like’ } \{\text{o} \cdot \text{ran}.j\text{o Ax}.t\text{Ix}, *\text{o} \cdot \text{ran}.j\text{Ax}.t\text{Ix}\}
\end{align*}
\]

There are several ways to solve this problem. I assume that \(-achtig\) has an underlying initial consonant like all (other) Class II suffixes. An obvious candidate for this would be the glottal stop which is also present on the phonetic surface. The advantage of this assumption is that the facts about \(-achtig\) now follow without any stipulation, because this suffix has the same phonological shape as \(-loos\) in all relevant respects:

\[
\begin{align*}
\text{rood} & \ {+} \ \{\text{ro.d}\} \ {?}\text{Ax}.t\text{Ix} \\
\text{roodachtig} & \ {+} \ {?}\text{Ax}.t\text{Ix}
\end{align*}
\]

Both candidates fare equally well with respect to these two constraints. Therefore, there are other constraints that decide between the two. \text{PR} \approx \text{LX} is one such constraint, but onset clusters of an obstruent followed by a glottal stop are also never found in Dutch. The constraint responsible for this, will naturally select candidate (b) in the table above.

We now turn our attention to the interaction with stress. ‘Class II’ suffixes are always stress-neutral. The stress pattern on their base is exactly the same as it would have been if the suffix were not attached. Furthermore, primary stress stays on the stem, even though the suffix may get a secondary stress:

\[
\begin{align*}
\text{a. } & \text{televisie ‘television’ } \{\text{tel} \cdot \text{v} \cdot \text{i} \text{zi}\} \\
\text{b. } & \text{televisie-achtig ‘television-like’ } \{\text{tel} \cdot \text{v} \cdot \text{i} \text{zi} \cdot \text{Axt} \text{Ax}\text{Ix}\}
\end{align*}
\]
‘Class I’ suffixes on the other hand are either ‘stress-attracting’ or ‘stress-bearing’. In the former case, stress falls on the stem, but on some other position than where it would be if the suffix were not attached (79). This is always a position closer to the suffix, hence the name ‘stress-attracting’. In the case of ‘stress-bearing’ suffixes, stress falls on the suffix (80).

(79)  
   a. eenvoud ‘simplicity’ [envaut]  
   b. eenvoud+ig ‘simple’ [envaudax]

(80)  
   a. respect ‘respect’ [respekt]  
   b. respectabel ‘respectable’ [respektbol]

It is widely agreed upon that the distinction between ‘stress-bearing’ and ‘stress-attracting’ suffixes can be derived from their respective phonological shapes. -abel ‘-able’ is stress-bearing because it is disyllabic, and furthermore, its final syllable contains a schwa; as a rule syllables immediately preceding schwa always attract main stress. -ig ‘-y’ on the other hand is monosyllabic and therefore is more likely to be stress attracting: as we have seen in the preceding section, final syllables only get stress in exceptional cases also in monomorphemic words. Most authors draw the conclusion from this that the stress in ‘Class I’ suffixed words is not in any essential way different from that in underived words. As far as stress is concerned, the boundaries between Class I suffixes and stems are invisible.

This observation was quite easily captured in the derivational framework of Lexical Phonology. In this framework, we can assume that the stress rules applied at the end of Class I, i.e. after Class I suffixation, but before Class II suffixation. If we assume furthermore that metrical structure is respected after it is built, we get the proper characterization of the facts: words derived at Class II receive a stress pattern much like a compound.

(81) | level: stress assignment | respect+abel | televisie-achtig |
    | Level I: stress assignment | [respektbol] | [tlevizi] |
    | Level II: compound stress | [telaviziachtix] |

Yet these same facts can be made to follow just as easily without the stipulation of lexical levels. The reason for this has already been sketched: Class I suffixes have an independent reason to cross morpheme boundaries. ONSET forces them to do this. Particularly relevant is also the constraint PR≈LX, the constraint which requires every morpheme boundary to correspond to a phonological word boundary. This constraint plays a decisive role in the derivation of a word like televisie-achtig. In forms like this, both the root and the affix get their own phonological word, because of PR≈LX. The stress on this form is therefore similar to that of a compound.

[ Booij & van Santen (1995); de Haas & Trommelen (1993); Booij (2002).]
2.3 Inflectional suffixes

The stress within each of the phonological words is determined by the constraints outlined in the previous section. However, we have seen that $\text{PR} \approx \text{LX}$ does interact with (is dominated by) $\text{ONSET}$. This was the core of our analysis why Class I suffixes are integrated into the syllabic structure of the stem. It can now also provide us with an explanation why they are integrated with the metrical structure of the stem:

\[
\begin{array}{c|c|c}
\text{word} & \text{stress} & \text{manner of stress acquisition} \\
\hline
\text{televizi} & + & \text{PR} \approx \text{LX} \\
\text{AxtIx} & (\text{televizi} & \text{AxtIx}) \\
\text{P} & (\text{televizi} & \text{AxtIx}) \\
\end{array}
\]

The difference in stress between Class I and Class II suffixes is therefore already explained: it is parasitic on their difference in segmental shape.

2.3 Inflectional suffixes

Inflectional suffixes in Dutch behave as ambiguous (or ‘paradoxical’) with respect to the criteria for Class I vs. Class II. On the one hand, they are insensitive to stress, just like Class II suffixes, but on the other hand they resyllabify, if they are vowel-initial, just like Class I suffixes.

I will discuss these two properties in turn. In the first place, we find in inflected forms we find exceptions to both generalisations in (??), repeated here for convenience:

\[
\begin{array}{c|c|c}
\text{word} & \text{stress} & \text{manner of stress acquisition} \\
\hline
\text{rEspEkt} & + & \text{ONSET} \\
\text{ab@l} & \text{PR} \approx \text{LX} \\
\text{Ekt} & (\text{rEspEkt} & \text{ab@l}) \\
\text{Ekt} & (\text{rEsp} & \text{Ekt}) \\
\text{ab@l} & \text{R} & \text{ab@l} \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{word} & \text{stress} & \text{manner of stress acquisition} \\
\hline
\text{rEspEkt} & + & \text{ONSET} \\
\text{ab@l} & \text{PR} \approx \text{LX} \\
\text{Ekt} & (\text{rEspEkt} & \text{ab@l}) \\
\text{Ekt} & (\text{rEsp} & \text{Ekt}) \\
\text{ab@l} & \text{R} & \text{ab@l} \\
\end{array}
\]

The difference in stress between Class I and Class II suffixes is therefore already explained: it is parasitic on their difference in segmental shape.

\[
\begin{array}{c|c|c}
\text{word} & \text{stress} & \text{manner of stress acquisition} \\
\hline
\text{rEspEkt} & + & \text{ONSET} \\
\text{ab@l} & \text{PR} \approx \text{LX} \\
\text{Ekt} & (\text{rEspEkt} & \text{ab@l}) \\
\text{Ekt} & (\text{rEsp} & \text{Ekt}) \\
\text{ab@l} & \text{R} & \text{ab@l} \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{word} & \text{stress} & \text{manner of stress acquisition} \\
\hline
\text{rEspEkt} & + & \text{ONSET} \\
\text{ab@l} & \text{PR} \approx \text{LX} \\
\text{Ekt} & (\text{rEspEkt} & \text{ab@l}) \\
\text{Ekt} & (\text{rEsp} & \text{Ekt}) \\
\text{ab@l} & \text{R} & \text{ab@l} \\
\end{array}
\]

*8* I will ignore here the class of derivational suffixes which are equally paradoxical. See van Oostendorp (2004).
a. [átłáso] ‘atlasses’ (*[átláso], from [átlas])

b. [ánváŋt] ‘start (3S)’ (*[ánváŋt], from [ánváŋ])

On the other hand, syllabification can see inflection. Inflected forms have a normal syllabification, as is visible if the inflection starts with a vowel. Notice that there is no final devoicing in (85). (Since there are no sonorant-initial inflectional suffixes, nothing can be said about them.)

(85) [lev@] (‘live (P L); from /lev/)

The traditional solution is (also) in this case, that suffigation of inflectional material will have applied after stress, but this leaves the question open why syllabification can reapply but stress cannot.

Yet another property of inflectional suffixes has to be mentioned: their segmental structure. Inflectional suffixes (as well as most function words in Dutch) are composed of segments in the following inventory:

(86) [t, s, n, r, ø]

If we assume that coronals are the least marked consonants, and that schwa is the least marked vowel of Dutch, we can see that these are the most unmarked stop, fricative, nasal, liquid and vowel of the language. Why would we find this tendency to unmarkedness?

The traditional solution does not have anything to say about this. How can we account for these curious properties? It seems a logical step to assume that the phonological differences between derivation and inflection should follow from the independently needed differences in morphological structure. Traditionally, it is assumed that derivational suffixes are morphological heads, because they determine the morphological category and other properties of the word. Inflectional suffixes, on the other hand, do not determine the category and therefore are not heads. We therefore assign the following morphological structures to these words:

(87) derivation inflection

We furthermore assume that phonological structure will try to mirror morphological structure as far as possible, for instance by way of ALIGN constraints. This means that the morphological difference between derivation

---

9The ideas presented here owe a lot to Revithiadou (1999), who applies kindred ideas to Greek.
and inflection will also be reflected in the phonology, which will aim for the following ‘optimal’ phonological word structures:

\[
\begin{align*}
\omega & \quad \omega \\
\text{derivation} & \quad \text{inflection} \\
\omega & \quad \omega \\
\text{half} & \quad \text{ling} & \quad \text{atlas} & \quad \text{en}
\end{align*}
\]

We have already seen, of course, that vowel-initial derivational suffixes get a different structure — one in which the suffix is completely incorporated —, for purely phonological reasons. However, consonant-initial derivational suffixes clearly aim for the structure represented here.

Inflectional suffixes thus will end up in phonologically adjoined positions, unlike derivational suffixes. Their special behaviour in the phonology follows from this. In the first place, if syllable structure behaves as outlined above, resyllabification is possible (\(\sigma\) takes highest constituent as domain).

Furthermore, let us assume that marked material needs to be licensed by being in a constituent; material adjoined to \(X\), does not really count as being dominated by \(X\) (\(\alpha\) is dominated by \(\beta\) iff \(\alpha\) is dominated by ever segment of \(\beta\)). If stress constituents need to be dominated by a word node \(\omega\), the stress behaviour follows.

Similarly, if marked segmental material needs to be dominated by \(\omega\), it follows that marked material cannot be in an adjoined position, hence that inflectional suffixes will not be able to carry marked segments. To make this more concrete, imagine we have constraints of the following type:

\[
\text{(89) \quad \text{WORD}(F): A phonological feature } F \text{ can only occur inside a word.}
\]

Assuming that features \(F\) are somehow arranged in an order of relative markedness (e.g. coronal \(\prec\) labial, velar) or some form of monovalency, we get the required result: only coronals are allowed in inflectional position.

One potential problem concerns the past tense suffix, which sometimes contains a voiced [d] (90). The feature [+voice] can hardly be considered unmarked.

\[
\text{(90) \quad } \text{ik lee[t]} \quad \text{‘I live’ - ik lee[v][d]a} \quad \text{‘I lived’}
\]

However, the past tense suffix only takes this shape if it follows a stem which has an underlyingly voiced obstruent, so that it can safely be assumed that this suffix is itself underlyingly \(-/t\alpha/\), and that voicing is shared with the preceding stem:
In this representation, the feature [+voice] is properly dominated by the verb (or the phonological word corresponding to it); the fact that it also occurs outside, is irrelevant for \textsc{Word([voice])}.

This also explains why in this case we have progressive assimilation of voicing. Usually voicing is in the opposite direction, e.g. in the nominalizing suffix -te:

(92) a. \textit{stijf-te} ‘stiffness’ from /steiv/ ‘stiff’
    b. \textit{stil-te} ‘silence’ from /stil/ ‘silent’

Notice that in this case, the suffix, which underlyingly has the same phonological shape, has a different (derivational) status. Apparently, the morphological origin of features is also important: morphological heads win.

3 Minimal morpheme expression

3.1 Introduction

One of the fundamental problems for constraint-based theories of phonology is the issue of opacity: a phonological process applies where it should not, or does not apply where it should, if we look at the phonological context on the surface. We find many cases of this in Dutch dialectology as well; a specific subclass is the topic of this class.

Leaving aside the question whether this level of abstractness is required for other cases, I argue that one class of cases of phonological opacity can be handled without stipulating an extra level of representation, but by taking into account the morphological structure of the forms in question. In particular, deleted segments sometimes still seem to influence the surface representation of morphologically complex words, since without this influence a whole morpheme would be lost. I argue that there is a principle of the following general shape:

(93) \textit{Phonological recoverability}. Every morpheme in the input should be represented in the phonological output.\textsuperscript{10}

\textsuperscript{10}Constraints which are similar to this in one way or another have been proposed among others by \cite{samek-lovdovici1993}, \cite{akinlabi1996}, \cite{granadesikan1997}, \cite{rose1997}, \cite{walker1998, walker2000}, \cite{piggott2000}, cf. \cite{kursu2001} for an overview.
A functional explanation to (93) is possible, if needed: if a morphologically complex form needs to be parsed, it is preferable to have cues in the phonological shape for every independent morpheme, but (93) can also be seen as a purely formal requirement on linguistic structure, perhaps a consequence of some more general principle of the architecture of the language faculty. In particular, it can be seen as an instance of what Jackendoff (1993) calls ‘correspondence rules’ between components of grammar; Jackendoff makes it clear that such rules satisfy a conceptual necessity under any view of the grammar.

It can be shown how a number of apparent cases of phonological opacity can be dealt with if we use this mechanism. My examples in the following are taken from the literature on various Dutch dialects. These have been fairly well-studied in the Dutch dialectological literature, but are not well-known outside of this tradition.

I have made a further restriction to inflectional morphology. The reason for this is that inflection usually is rather ‘weak’ as we have seen in the preceding class. If any morpheme ever is a likely candidate for violating (93), it most likely is an inflectional morpheme. Furthermore, we have a relatively clear view of the internal morphological structure of inflectional elements (which consist of purely ‘formal’ features only), whereas this is much less the case for derivational affixes, in which some amount of lexical semantics is also involved.

The structure of the argumentation will be the same in each example. An inflectional morpheme is phonologically weak in the way just outlined and therefore bound to be deleted. At the same time, if it would be present, it would either trigger or block a process of assimilation. In order to satisfy the requirement in (93), the deletion of the morpheme is not complete; the constituent of the original segment which is necessary to participate either positively or negatively in the assimilation process is left behind as a trace. For example, in Hellendoorn Dutch, an otherwise active process of progressive nasal consonant assimilation seems to be blocked in the past tense (in the cases below, the plural suffix may be assumed to be syllabic /n/; the orthographic examples represent Standard Dutch):

(94)  a. *werken* ‘(to) work’ [werkə]  
      b. *werkten* ‘(we) worked’ [werkə]  
      c. *hopen* ‘(to) hope’ [hopə]  
      d. *hoopten* ‘(we) hoped’ [hopə]

As can be seen from the orthography, and as will become evident if we study other instances in Hellendoorn Dutch, the imperfective suffix underlyingly contains at least a coronal obstruent /t/. We can now analyze this as a case of rule opacity: first we have an assimilation rule, and afterwards a rule of t deletion, obscuring the original environment of assimilation.
An alternative approach is to assume that /t/ is not deleted fully, but leaves behind a trace, in the form of the feature [coronal], which is then realized on the nasal consonant. The reason for this could be a general requirement that linguistic structure should be visible and expressed, i.e. the principle in (93). The consequences of this approach are explored in this article. The discussion will be embedded within Optimality Theory, currently the most popular theory of input-output mapping, but (93) is virtually theory-independent and its effects could be couched within other frameworks as well.

We could wonder whether many apparent cases of opacity do not actually disappear if we assume a somewhat more sophisticated view of phonological representations, taking into account the literature on prosodic organization, autosegmental structure and the interaction with morphological and syntactic boundaries. One representation, which is enriched by independent necessary elements, may then do the work of two poorer representations.

In Antwerp Dutch, for instance, we have a process velarising a nasal consonant in coda position. We also have a process shortening a vowel before the resulting velar nasal (Taeldeman, 1982). Both processes can be seen at work in the following examples:

\[(95)\]
\[\text{a. grune ‘green’ } [\text{gyr}:\text{na}] \sim \text{gruun ‘green’ } [\text{gyr}:\text{y}]^{11} \]
\[\text{b. schoenen ‘shoes’ } [\text{sxu}:\text{no}] \sim \text{schoen ‘shoe’ } [\text{sxu}:\text{n}]\]

Interestingly, the velarisation process only applies to words with underlingly long vowels (Antwerp Dutch presumably differs from Standard Dutch in having really long vowels, although this is not absolutely crucial to us), and not by words which have short vowels already underlingly:

\[(96)\]
\[\text{a. kin ‘chin’ } [\text{kin}] \sim \text{tien ‘ten’ } [\text{ti}:\text{n}] \]
\[\text{b. zon ‘sun’ } [\text{zon}] \sim \text{zoon ‘son’ } [\text{zon}]\]

In order to describe this, we could write the following rules (following Taeldeman, 1982):

\[(97)\]
\[\text{a. } n \rightarrow \text{y} / V:\text{---} \]
\[\text{b. } V: \rightarrow \text{V} / \text{--- y}\]

When applied to the different inputs in the right order, these rules will yield the correct results. While strictly speaking the rule in (97a) is opaque, this is only so because it has a rather unnatural shape: it is very uncommon in

\[^{11}\text{Apart from some cases of optional schwa deletion, the form without schwa can only be used in the singular neuter of adjectives in indefinite noun phrases. The form with schwa can be used in all other inflections (plural or non-neuter, or both).}\]
languages of the world for velar nasals to only show up after long vowels. As a matter of fact, most variants of Dutch (and Germanic) allow the velar nasal to occur only after short vowels. This is even true for Antwerp Dutch, at the surface; rule (97b) is responsible for that. A much more natural rule would therefore be the one in (98):

(98) \( n \rightarrow \eta \ / \ V \) ___

Yet this process would be very opaque indeed, since we obviously find cases where underlying /\( n \)/ did not turn into a velar. We would therefore have to distinguish between underlingly short and shortened vowels: the process seems so opaque that even an analysis based on rules (with arbitrarily many intermediate representations) cannot deal with it satisfactorily.

On closer inspection, there is evidence that the velar nasal, different from the other nasal consonants, is restricted to the coda position in many varieties of Dutch (Trommelen, 1982; van Oostendorp, 2001; van der Torre, 2003). This could explain, for one thing, the fact that velar nasals can only occur after lax vowels, given the fact that only these can occur in closed syllables. A velar nasal after a tense vowel would then be prohibited, because it could only occur in an onset:

\[
\begin{array}{cccc}
\text{bang} & \text{‘afraid’} & \text{ban} & \text{‘ban’} & \text{baan} & \text{‘ban’} & \ast \text{baang} \\
\mu & \mu & \mu & \mu & \mu & \mu & \mu \\
\text{b} & \text{a} & \eta & \text{b} & \text{a} & \text{n} & \text{b} & \text{a} & \text{n} & \text{b} & \text{a} & \eta
\end{array}
\]  

(99)

There undoubtedly are other ways to capture the same intuition: that both long vowels and velar nasals need space in the syllable and that if we would have both, this would be too much. Under the one chosen here, the opacity of the Antwerp Dutch velarisation process disappears, if we assume that there is a strong faithfulness requirement on the number of mora’s in this dialect — or in rule-based terms, that we are not allowed to insert any mora’s in the underlying representation. A change from /\( ti:n \)/ — two underlying mora’s attached to the vowel — to [\( \text{tiij} \)] (two surface mora’s, one for the vowel and one for the nasal) is then allowed, but a change from /\( ki:n \)/ (one underlying mora attached to the vowel) to [\( \text{kiij} \)] (two surface mora’s) is not. If we allow ourselves to introduce a few ad hoc constraints to make things work technically, an OT analysis might then run along the following lines. We need the constraints in (100), and the ranking in (101) in order to get the tableau in (102):

\footnote{It is sometimes assumed that the reason for this restriction is that velar nasals underly-ingly are /\( \text{ng} \)/ or /\( \text{ng} \)/, just like this is assumed for English. The Antwerp facts actually show that this analysis cannot be correct: here we find the same restriction but there is absolutely no possibility for postulating an underlying obstruent.}
3.2 Nasal assimilation in Hellendoorn Dutch past tense

As outlined above, Hellendoorn Dutch — a dialect spoken in the northeastern parts of the Netherlands —, like many other languages in the world, displays a process of nasal assimilation. Interestingly, the process works from right to left as well as from left to right. The following facts are all from Nijen Twilhaar [1990], the orthography again is Standard Dutch:

---

**3.2 Nasal assimilation in Hellendoorn Dutch past tense**

More in general it seems to be the case that next to a more sophisticated view of phonological structure, also a more precise view of the interaction between phonology and morphology, and in particular of the ‘visibility’ of morphology for phonology, may help to make many apparent examples of phonological opacity actually vanish. Constraints to not uniquely refer to segments, arranged in a one-dimensional string; what we actually have is a more fine-grained, multidimensional vision referring to the internal structure of these segments, and to the relations between them.13

3.2. Nasal assimilation in Hellendoorn Dutch past tense

Nasal assimilation in Hellendoorn Dutch has some interesting properties. Examples (103a-c) show that a (syllabic) nasal assimilates to a preceding obstruent. In contradistinction to the first analysis in (94), the plural suffix is represented here as underlyingly /@n/. I will return to this assumption below. For now it suffices to see (103d) that the indefinite article, which unquestionably has a schwa underlyingly (because this schwa surfaces e.g. if an indefinite nominal phrase occurs at the beginning of a sentence), displays the same behaviour. (103e) shows that nasals in onset position are not affected by the process, and (103f) that in certain cases assimilation is regressive, to a following consonant.

The key facts are the ones in (104):

<table>
<thead>
<tr>
<th></th>
<th>orthography</th>
<th>underlying</th>
<th>surface</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>lopen</td>
<td>lop@n</td>
<td>lop@n'</td>
<td>‘to walk’</td>
</tr>
<tr>
<td>b</td>
<td>weten</td>
<td>wet@n</td>
<td>wet@n'</td>
<td>‘to know’</td>
</tr>
<tr>
<td>c</td>
<td>p@kken</td>
<td>p@k@n</td>
<td>p@k@n'</td>
<td>‘to grab’</td>
</tr>
<tr>
<td>d</td>
<td>loop een</td>
<td>lop an</td>
<td>lop@n'</td>
<td>‘(I) walk a (mile)’</td>
</tr>
<tr>
<td>e</td>
<td>rampnacht</td>
<td>rampn@xt</td>
<td>rampn@xt</td>
<td>‘disastrous night’</td>
</tr>
<tr>
<td>f</td>
<td>loop een keer</td>
<td>lop an ker</td>
<td>lop@k@k</td>
<td>‘(I) walk one time’</td>
</tr>
<tr>
<td>g</td>
<td>koop een koek</td>
<td>kop an kuk</td>
<td>kop@k@k</td>
<td>‘(I) buy a cake’</td>
</tr>
<tr>
<td>h</td>
<td>kopen koeken</td>
<td>kop an koeken</td>
<td>kop@k@k</td>
<td>‘(we) buy cakes’</td>
</tr>
</tbody>
</table>

This is a case of opacity because within a rule-based framework, we could state two rules (disregarding regressive assimilation), one of progressive nasal assimilation, and another one of /t/ deletion (the following is based on Nijen Twilhaar, 1990):

(105) a. \( t \) deletion: \( t \rightarrow \emptyset / C \)  

b. progressive assimilation (PA): \( \text{[nasal]} \)

PA is rendered opaque by \( t \) deletion (schwa deletion is implied to be proceeding the processes described here):\(^{14}\)

\(^{14}\)Interestingly, this is a case of opacity either of type (i) or of type (ii) in terms of Kiparsky’s definition in (?), depending on how we look at it. We have surfacing CAD in the sense that it looks as if the nasal has not been subject to assimilation even though the context is present; we have XBY in the sense that it looks as if the nasal has been subject to assimilation to a segment...
3.2. Nasal assimilation in Hellendoorn Dutch past tense

(106) /stop+t+an/ /zert+t+an/ /pak+t+an/

schwa deletion stOptn zetttn pakttn
PA stOptn zetttn pakttn
t deletion stOptn zetttn pakttn

It is fairly easy to set up an analysis of the non-opaque facts in (103). Again, we use a few constraints which may not be hallmarks of theoretical sophistication, but which give the required results.\(^{15}\)

(107) a. ASSIMILATE: A coda nasal and an adjacent obstruent should have the same place of articulation.

b. \(^{*}\)CCC: Clusters of three consonants are not allowed.

c. FAITH(PLACE): Input place features should surface.

(108) /pak+t+an/

\(^{*}\)CCC ASSIMILATE FAITH(PLACE)

\(\text{pakn} \) *! W
\(\text{patn} \) *! W

Hellendoorn differs from other languages displaying faithfulness of place features in that even after the consonant deletion, another obstruent stays present that could still enforce assimilation. Therefore, the opaque cases here cannot be dealt with without additional means.\(^{16}\)

(109) /pak+t+an/

\(^{*}\)CCC ASSIMILATE FAITH(PLACE)

\(\text{xpakn} \) L
\(\text{patn} \) *! W

What we need to express, here, is the idea that the nasal gets its feature from the underlying past tense suffix. One way of doing this, is by formalizing the following constraint:

which is no longer there. I have chosen the second possibility here. There would be ways to test which of these two theories is correct, if we would be able to find e.g. cases where the deleted consonant is non-coronal.

\(^{15}\)It would certainly be possible to give more sophisticated analyses using more elegant constraints, but these would require more different constraints, and the point would remain the same: an extra faithfulness constraint is necessary to understand the exceptional behaviour of past tense forms.

\(^{16}\)It is assumed here that deletion of /k/ or /n/ is not an option for satisfying \(^{*}\)CCC. As we will see below, /t/ is particularly prone to deletion in dialects of Dutch, but other consonants cannot be deleted. Exploring the reasons for this is beyond the scope of this article.
3.2. Nasal assimilation in Hellendoorn Dutch past tense

(110) **EXPRESS-[F]**: The morphological feature F should be expressed in the phonological surface.

(Some phonological feature connected to the input expression of F should be present in the output.)

This is a special type of faithfulness constraint, basically stating that it is not allowed to delete a morpheme fully. An instance of this general constraint scheme could now be **EXPRESS-TENSE**: some part of the past tense suffix /-an/ should be expressed in the output. Adding this constraint to our tableau gives us the desired result:

![Tableau](image)

How exactly does **EXPRESS-TENSE** work? The phonological input is a /t/, i.e. a feature bundle like the following:

(112) $\omega_{\text{tense}}$

$$[\text{coronal}]_{\text{tense}} \quad [-\text{sonorant}]_{\text{tense}}$$

Every morpheme consists of a number of feature bundles, connected to a root note, and/or a timing slot. I marked this by adding a subscript to every individual element. Seen in this notation, **EXPRESS-TENSE** states that the output should contain at least one element which has this subscript.

Until this point, we have silently assumed that the plural suffix /-an/ has a schwa underlingly. We can find some arguments for this in Nijen Twilhaar (1990). Most convincing perhaps is the argument that we also find monomorphemic nouns ending in a syllabic nasal, and nouns ending in /-an/, but no nouns ending in /-oan/. This shows at least that a productive process of schwa deletion before n is going on. Furthermore, the schwa sometimes surfaces, viz. in very formal styles of speech (Nijen Twilhaar [1990]:165); these are styles where typically the surface form is closer to the underlying representation (van Oostendorp, 1997).

There are various reasons why schwa should not surface; being phonologically and phonetically empty, it seems a less desirable nucleus, etc. There will thus be a constraint *o or something more motivated but to the same effect (cf. van Oostendorp, 2000, for fuller detail).

We can distinguish between three groups of dialects of Dutch (van Hout & van der Velde, 2000). In some, schwa is deleted, and in others /n/ is deleted under various circumstances; in line with the previous discussion, this could be formalized as a constraint *n. The third variety is one in which
3.2. Nasal assimilation in Hellendoorn Dutch past tense

neither schwa nor /n/ is deleted. Crucially absent are those dialects in which both schwa and /n/ are deleted. This may be seen as an indication for the high level of activity of recoverability, formalized in this case as a constraint EXPRESS-PLURAL. This constraint would then dominate at least one of *a and *n.

These facts are not unique for Hellendoorn Dutch; we find very similar phenomena even in typologically unrelated languages. The interaction between nasal assimilation and consonant deletion from the Ojibwa dialect of Odawa [Piggott 1974; Kaye 1974] show a very similar pattern:

(113) Underlying /takossin-k/ ‘he arrives’
Assimilation takO$ink
Deletion takO$in

These facts are clearly very similar to those of Hellendoorn in the relevant respects. Viewed from a purely segmental point of view, Assimilation and Deletion are in an opaque (‘counterbleeding’) order. Yet if we consider the possibility that the place feature [velar] is on an independent plane, and that it can be retained even after deletion of the segment /k/, these facts follow. Also in this case, it appears that /k/ is an independent morpheme, having a conditional meaning. If we assume that this conditional morpheme has to be retained at the surface somehow, the feature [velar] would then show up on the nasal as a trace of this morpheme in order to satisfy EXPRESS-[cond]. All in all, we would get an analysis such as the following:

(114) a. EXPRESS-[cond]: The conditional morpheme should be expressed at the surface
b. *CC: Consonant clusters are not allowed.

The tableaux of Hellendoorn and Ojibwa thus become strikingly similar. Again, there is no need to rank most of the relevant constraints in order to get the difference between conditionalis and realis forms:

(115)

| *takoj|j|  |
|----|----|----|----|
| /takoss|in/ | C.C. | EXPRESS-COND | ASSIMilate | FAITH[PLACE] |
| tako$|in| |  |  | *W |

(116)

| *takoj|j|  |
|----|----|----|----|
| /takoss|in+k/ | C.C. | EXPRESS-COND | ASSIMilate | FAITH[PLACE] |
| tako$|ink| |  |  | *W |
| tako$|ink| |  |  | L |
| tako$|ink| |  |  | *W |
| tako$|ink| |  |  | L |

As a matter of fact, there thus is no opacity, or any problem for surface-based phonology within Hellendoorn Dutch or Ojibwa at all, given the fairly standard assumption that place or articulation features can exist independently
of their segments — an assumption that was not available in the work of Kiparsky and Kaye just referred to above.

3.3 Voicing Assimilation in Flanders and Brabant

The next example which deserves discussion is widespread in the Dutch-speaking parts of Belgium (at least in Flanders and Brabant). In this case a process of voicing assimilation interacts with the deletion of word final /t/, which is the phonological shape of the third person singular verbal inflection (Taeldeman 1982) in a way that may be considered opaque:

\begin{align*}
(117) & \quad \text{a. } [-\text{sonorant}, +\text{continuant}] \rightarrow [-\text{voice}] / [-\text{sonorant}] \_ \\
& \quad \text{b. } t_{3,sg} \rightarrow \emptyset / \_ \_ \_ \# C
\end{align*}

\begin{align*}
(118) & \quad \text{hij doe}/t \ v/eel \ 'he does a lot'
& \quad \text{a. } \text{hij doe}/t \ v/eel
& \quad \text{b. } \text{hij doe}[f]eel
\end{align*}

Voicing assimilation in Dutch is a well-known and widespread phenomenon (cf. Lombardi 1999; van der Torre & van de Weijer in press, for various analyses within the OT framework). Lombardi (1999, p. 277) analyses it in terms of the following constraint:

\begin{align*}
(119) & \quad \text{AGREE: Obstruent clusters should agree in voicing.}
\end{align*}

AGREE is of course very similar in form and spirit to the constraint ASSIMILATE, which we used above to describe nasal assimilation. Under this formulation, the process in question becomes unmistakably opaque, quite independent where we rank AGREE (if $^*CC \gg \text{FAITH(VOICE)}$):

\begin{align*}
(120) & \quad \text{\begin{tabular}{|c|c|c|}
| [f] & | & $^*!$ \\
\hline
\text{hij doe}/t \ v/eel & $^*CC$ & \text{AGREE} \text{ FAITH(VOICE)} \\
\text{[tv]} & $^*!$ \text{W} & $^*W$ \_ L \\
\text{[tt]} & $^*!$ \text{W} & * \\
\hline
\text{[v]} & L &
\end{tabular}}
\end{align*}

In order for a solution along the lines of EXPRESS-[F] (in this case: EXPRESS-[3PS]) to work, we need to know what exactly is the phonological element that expresses the inflectional suffix in this case. Lombardi (1999) employs a theory of laryngeal features in which [voice] is monovalent, i.e. there is no phonologically active [-voice]. This means that the underlying and surface representations are schematically as follows (representing [v] as a /f/ with attached [voice]):
The underlying representation of the [3sg] suffix seems to have disappeared without leaving a trace; there is nothing in the surface form to represent it, given the plausible assumption that absence of a feature cannot act as a representative.

Wetzels & Mascaro (2001) and various other authors have argued on independent grounds that there are empirical arguments to assume that [-voice] should be assumed to be phonologically present. In that case, the inflectional suffix does indeed leave a trace at the surface representation, viz. the feature [-voice], realized on the [f]:

(122)  

Wetzels & Mascaro (2001) and various other authors have argued on independent grounds that there are empirical arguments to assume that [-voice] should be assumed to be phonologically present. In that case, the inflectional suffix does indeed leave a trace at the surface representation, viz. the feature [-voice], realized on the [f]:

(122)  

In van Oostendorp (2002), on the other hand, it is argued that the difference between voiced and voiceless fricatives phonologically behaves like a length distinction in many Westgermanic dialects. Intervocally, at least, ‘voiced’ fricatives are short, and ‘voiceless’ fricatives are long. This explains, among other things, why in many of these dialects we find voiced fricatives after tense or long vowels and voiceless fricatives after lax or short vowels (the following examples are from Dutch):

(123)  

These facts are easily explained if — given our analysis so far — tense vowels occur in open syllables and lax vowels occur in closed syllables, and voiceless fricatives are ambisyllabic (so that they close the syllable):

(124)
(124c) has a tense vowel in a closed syllable, and (124d) a lax vowel in an open syllable; both are excluded by CONNECT (??), whereas the structures in (124a) and (124b) are correctly allowed.

There is some empirical support for this assumption in the work of Ernestus (2000, p. 177). Based on a corpus of spontaneous (Standard Dutch) speech, Ernestus notes that clusters of fricatives of the same place of articulation arise when a word-final fricative is followed by a word-initial one. These clusters are generally realized with a duration that is shorter than the duration of two segments (...). In what follows, clusters consisting of two segments with the same manner and place of articulation will be referred to as geminates. [...] The problem is that fricative geminates are always realized as voiceless, independently of their context, exact duration, etc.

From this we can thus at least conclude that longer fricatives are always voiceless.

If this analysis is correct, it could be the position of the /t/ that is retained after the disappearance of the segment:

```
(125) underlying               surface
      h e i d u t₃,e l            h e i d u ♦, e l
      [labial]                   [labial]
```

It thus is the phonotactic position (represented here as a dot, since it does not matter whether this is a mora, a root node, or something else), that could be seen as the trace of the suffix, necessary to satisfy EXPRESS-3SG.

```
(126)  ε= [t]    |   |   |  *  |
       hij doe/t v/eel  *CC  EXPRESS-3SG  AGREE  FAITH(VOICE)
       [tf]     *! W                *  |
       [tv]     *! W                * W  L
       [v]       *! W                * W  L
```

Notice that the tableau is very similar to the ones given for Hellendoorn and Ojibwa above: the morphological faithfulness constraint EXPRESS-3SG outranks the phonological faithfulness constraint FAITH-voice.

Taeldeman (1982) asserts that the opacity effect attested in the dialects just discussed is typical for deleted /t/ as a morpheme. Similar effects can however be found in the literature purely internal to the phonology in other dialects of Dutch. E.g. in Wilsum Dutch, the final /t/ of function words such
as net (just), met (with) and det (that) can disappear, but still have the effect of devoicing the following /t/. We thus get examples such as the following (Spa, 2000, 46):

(127) a. ne/t z/o dudelijk > ne/slo dudelijk ‘just as clear’
    b. me/t z/i/en ome > me/sli/en ome ‘with his uncle’
    c. as ie de/t v/raogen > as ie de/t]raogen ‘if he that asks=if he asks that’

In this case, the relevant type of faithfulness cannot be morphological, since the function words in question are expressed by an onset consonant and a vowel; EXPRESS constraints are thus satisfied in any case. It could be argued that we thus have an instance of faithfulness to the segmental position, usually formalized as a DEp constraint (every segment in the input should have a correspondent in the output). The correspondent of /t/ will be filled by material from the fricative (which means that a lower-ranking IDENTITY-constraint will be violated).

(128)

<table>
<thead>
<tr>
<th></th>
<th>*CC</th>
<th>DEp</th>
<th>AGREE</th>
<th>FAITH(VOICE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ne/t z/o dudelijk</td>
<td>*CC</td>
<td>DEp</td>
<td>AGREE</td>
<td>FAITH(VOICE)</td>
</tr>
<tr>
<td>[ts]</td>
<td>*! W</td>
<td>L</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[tz]</td>
<td>*! W</td>
<td>* W</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

The situation is somewhat more complicated, because we need to explain why this only involves the final /t/ of function words. It has been argued that the final /t/ of words such as these is also in some abstract sense a suffix (Vanacker, 1949): the /t/ shows other types of deviant behaviour in other contexts as well. Taeldeman & Schutter (1986, 114) propose that there is a hierarchy of positions where devoicing after deleted /t/’s may occur:

i. verbal inflection
ii. the ‘small words’ daT ‘that’, waT ‘what’, nieT ‘not’ and (sometimes) meT ‘with’
iii. frequent adjectives of the type V:d (e.g. goed ‘good’, kwaad ‘angry/bad’, koud ‘cold’, dood ‘dead’)

This hierarchy seems to reflect the straightforwardedness of the inflectional nature of the /t/ in question. The hierarchy is based on the fact that in some dialects (the ones just discussed) we only find this effect in environments (i), in other dialects (the one spoken in Bruges) we find them in (i) and (ii), and yet in others (e.g. Ghent) we find them in (i), (ii) and (iii).
The question then remains why only frequent adjectives of a specific shape participate in Ghent. The frequency effect might be attributable to the fact that there tends to be more deletion in frequent words than in less frequent words in general (Goeman, 1999). It is not exactly clear to me why there should be a preference for adjectives ending in long vowels plus /d/. Yet one aspect is of particular importance here: the fact that in this case an analysis in which an underlying feature [-voice] would be the trace of the deleted segment cannot work. This gives indirect support for the analysis presented above in which it is the position of the coronal stop that surfaces, filled with the material of the fricative, which thereby lengthens and thus remains voiceless.

### 3.4 Remnants of the first person singular morpheme

A relatively well-known case of opacity within Dutch dialectology is provided by Aalst Dutch. At least according to the very careful phonetic study of Colinet (1896), this dialect used to display an interesting opaque application of nasal assimilation: the disappearing element (an inflectional schwa) did not trigger assimilation (as was the case in the previous examples of voicing assimilation), but rather to block it (Colinet, 1896; Taeldeman, 1980). (Taeldeman (2002) reports that some speakers still had the phenomenon in fieldwork in the second half of the twentieth century.) If an inflectional schwa (in the case at hand a schwa which expressed adjectival agreement) disappeared before a nasal, the nasal did not assimilate to the consonant following it on the surface, in spite of the fact that the environment for assimilation seems to be present:

\[
\begin{align*}
\text{schoon/ventje} & \quad \text{'handsome guy'} & \text{schoon/+/-/vrouw} & \quad \text{'beautiful woman'} \\
\text{schoon/ventje assimilation} & \quad \text{schoon/+/-/vrouw d.n.a.} \\
\text{schoon/ventje d.n.a.} & \quad \text{schoon/+/-/vrouw schwa deletion}
\end{align*}
\]

(130)

It is hard to see what the ‘trace’ of the adjectival inflection could be that would be necessary to satisfy EXPRESS-[Agr] (‘Agr’ standing for whatever the morphological features of overt adjectival inflection are). The reason for this is that the schwa seems to have gone on the surface completely, and there is nothing in e.g. the segmental make-up final [n] of schoon that could be seen as a trace of the existence of this segment:
3.4. Remnants of the first person singular morpheme

How can we explain these facts? The crucial observation is that nasal consonants only assimilate in dialects of Dutch if they are not in syllable onset. The conclusion therefore is that the trace of the agreement morpheme that is necessary in order to derive the opacity effect, is the syllable structure. We could suppose, for instance, that the inflectional schwa has a mora underlyingly. Under this view, it is the mora then that serves as the trace of adjectival inflection. The mora then necessarily projects an (empty-headed) syllable. Notice that this should mean that this morphologically sponsored empty head should have a different status from a purely phonological one (Scheer, 2004).

A similar account may be able to explain a quite spectacular examples, cf. the lack of final devoicing in Tilligte Dutch (Goeman, 1999, p. 216). Even though this dialect displays the effects of syllable-final devoicing elsewhere pervasively, we find forms such as ik geleuv ‘I believe’ (or ik geleuw) in the first person singular (Schoemans & van Oostendorp, 2004; van Oostendorp, 2005b). Importantly, Goeman (1999) notes that in neighbouring dialects we find a schwa serving as an overt first person singular suffix in the neighbouring dialects where the suffix has not yet been lost. The analysis here could be exactly the same as for Aalst Dutch:

(133)  \[
\text{underlying} \quad \sigma \\
\chi \ \omega \ \hat{\sigma} \quad \sigma
\]

(134)  \[
\text{underlying} \quad \mu_{agr} \\
\chi \ \omega \ \hat{\sigma} \quad \sigma
\]

(135)  \[
\text{underlying} \quad \mu_{agr} \\
\chi \ \omega \ \hat{\sigma} \quad \sigma
\]

(136)  \[
\text{underlying} \quad \mu_{agr} \\
\chi \ \omega \ \hat{\sigma} \quad \sigma
\]

(137)  \[
\text{underlying} \quad \mu_{agr} \\
\chi \ \omega \ \hat{\sigma} \quad \sigma
\]

(Goeman, 1999, 216-217) lists a large number of dialects where this phenomenon may be found; furthermore such dialects can be found in quite a large part of the Dutch language area. The reason Goeman gives for this, is a historical one: the first person singular schwa has been deleted ‘recently’ and therefore the final devoicing has not yet taken place. We could say that this
3.4. Remnants of the first person singular morpheme

statement depends on the opacity of diachronic language change: the final devoicing process proceeds as if the historical ending were still there.\footnote{Taeldeman & Schutter (1986); De Vriendt & Goyvaerts (1989); Goeman (1999); van Bree (2003); Schoemans & van Oostendorp (2004); van Oostendorp (in press).}

It is of interest that once again fricatives are the main focus of this exceptional behaviour, as is to be expected if we assume that fricative voicing is primarily an issue of syllable positions and those positions can be used to express morphological structure.

Other phonological processes may also be influenced by this vocalic position and the onset it licenses. In Brussels Dutch (De Vriendt & Goyvaerts \cite{DeVriendtGoyvaerts1989}) we see that various phonological processes act as if the first person singular still ends in a vowel. For instance, words in this dialect are not allowed to end in a velar nasal. Words which have such a segment underlyingly, develop a [k] at the end by some process of /k/ insertion:

\begin{enumerate}
\item \textit{k-insertion‘}  
\begin{align*}
\text{[puli[k]‘eel’ & \quad [puli[ø]‘eels’} \\
\text{[ɣa[ø]‘corridor’ & \quad [ɣa[ø]‘corridors’}
\end{align*}
\end{enumerate}

We can see that the /k/ is inserted here at the end of the word since it does not occur in the plural forms, before a schwa (there is a difference here with a form such as \textit{plank-planken ‘plank(s)’} which do have an underlying /k/).\footnote{Alternatively, one might argue that these words end e.g. in /nø]/ underlyingly, which devoices at the end. In this case, we need to say that the /ø/ is deleted in the first person singular, just as it is deleted before a schwa. The puzzle for the phonology-morphology interface stays the same.}\footnote{The first person singular preterite did not have a schwa ending and therefore is irrelevant from our perspective.}

Yet velar nasals can be found at the end of verbs in the first person singular (present):

\begin{enumerate}
\item \textit{ik hang ‘I am hanging’} [ikø] \\
\item \textit{ik zing ‘I am singing’} [iksø]
\end{enumerate}

The behaviour of this form could be explained in various ways, depending on one’s analysis of the behaviour of the velar nasal. One could state for instance, that a velar nasal is not allowed to occur as the last segment of the syllable. In the first person singular, this condition would not apply, since there is an (empty) syllable head following the velar nasal. The segment therefore would occur in the onset of such a syllable, and there would be no need to insert a /k/.

In the same dialect, we find paradigms such as the following:

\begin{enumerate}
\item \textit{kleden ‘to dress’} [kleja] \\
\item \textit{hijkleedt hem ‘he dresses himself’} [a+klit+am] (shortening)
\end{enumerate}
Tonal dialects

4.1 Introduction

Limburg dialects of Dutch have two distinctive tonal contours on syllables with primary stress. These tones are traditionally called Schleifton (‘dragging

There have been several proposals in the literature that word-final consonants are onsets rather than coda’s (cf. Piggott 2002 for an overview). Most of these do not differentiate between morphological contexts: all words are supposed to end in a consonant. Such proposals cannot differentiate between the two instances of open in Standard Dutch, and the two instances of geloof ‘belief’ (the verbal form which ends in a voiced consonant and the nominal form which is subject to final devoicing) in Twente, or at least they have to find another way to do so.
tone’) and Stosston (‘bumping tone’), but here we will use the terms ‘level high tone’ and ‘falling tone’. The tones fall on the stressed syllable in the word, and serve to distinguish between minimal pairs. The following examples are from the Maasbracht dialect:

\[
\begin{array}{lll}
\text{failing tone} & \text{level high tone} \\
\text{mǔñ ‘minus’} & \text{mǔñ ‘vile’} \\
\text{dáéñ ‘fir’} & \text{dáéñ ‘then’} \\
\text{klááñ ‘trap’} & \text{klááñ ‘hardly’} \\
\text{bũñ ‘bee’} & \text{bíñ ‘with’} \\
\text{zéñ ‘side’} & \text{zéñ ‘she’} \\
\text{púñp ‘to squeak’} & \text{púñp ‘pipe’} \\
\end{array}
\]

The distinction between these two tones is also used to in inflectional morphology, e.g. to differentiate between neuter and feminine forms of adjectives (139a); if the neuter is level high (wíñs), the feminine has a falling tone (wíñs). If the neuter itself has a falling tone, nothing happens to the feminine, which still has a falling tone (139b).

\[
\begin{array}{lll}
\text{neuter} & \text{feminine} & \text{masculine} \\
\text{a.} & \text{wíñs} & \text{wíñs} & \text{wíñs} & \text{‘wise’} \\
& \text{dóñuf} & \text{dóñuf} & \text{dóñuf} & \text{‘deaf’} \\
& \text{láñm} & \text{láñm} & \text{láñm} & \text{‘lame’} \\
\text{b.} & \text{káñlm} & \text{káñlm} & \text{káñlm} & \text{‘calm’} \\
& \text{kéñén} & \text{kéñén} & \text{kéñén} & \text{‘small’} \\
\end{array}
\]

The only distinction between the neuter form and the feminine form thus is one of tone. Given the fact that neuter adjectives can have both falling and level tones, depending on lexical specification, it is reasonable as well as customary to assume that this form of the adjective represents the ‘underlying’ tonal distinction.

These facts have been taken by some analysts (notably as evidence for the relevance of paradigmatic relations within phonology: the tones in (139a) would switch because in this way an opposition within the paradigm would be maintained (and higher-ranking markedness constraints would make such a switch impossible in cases such as (139b)).

We defend what could be called a more ‘traditional’ approach to these facts, assuming a combinatorial view of morphology in which all alternatives are due to the fact that one word consists of a different combination of morphemes than another word. There are no ‘paradigms’ in this view, only

\[21\text{This class is based on} van Oostendorp 2006.\]
\[22\text{Hermans} 1994.\]
\[23\text{A similar distinction is made in the realm of nouns, where singular nouns may carry a level tone, while the corresponding plurals have a falling tone.}\]
4.2 Tones and adjectival inflection: data

The phonology of tones

Limburg Dutch dialects, like the neighbouring Rhineland German dialects, are well-known for their use of lexical tone. There is quite some dialectal variation as to the phonetic realisation of these tones, but as far as is known, this does not really affect the phonology: the split between falling tone and level high tone is common to all dialects in this area.

In order to understand the interface between the phonology and the morphology, it is first necessary to understand the phonological identity of the so-called falling tone and the so-called level high tone. The following two pictures represent the F0 values for these two tones (for a speaker from the Roermond dialect, very close to Maasbracht24):

(140) falling tone level high tone

The ‘falling’ tone is characterised by a clear downward movement; the ‘level high’ tone also moves slightly downward, but then goes up again towards the end. There are several ways to translate this into the phonology, but many analysts have converged on the following (see Gussenhoven, 2004, for an authoritative overview):

(141) falling tone level high tone

\[
\begin{align*}
\text{HL} & & \text{HL}\text{H} \\
/\mu/ /\iota/ & & /\iota/ /\iota/ \\
m / h & & m / h
\end{align*}
\]

24The data were analysed with the Praat programme: http://www.praat.org/ The data are almost identical to those presented in Gussenhoven (2000).
Dialectology

The Limburg dialects are spoken in Dutch and Flemish provinces which are both called ‘Limburg’. Like most dialects in Europe, they are under a strong pressure of convergence to the standard language, in this case to Standard Dutch, but maybe to a slightly lesser extent than in some other areas in this particular corner of Europe (Kroon & Vallen 2004). The area is on the periphery of the Dutch-speaking area, neighbouring both German and French dialects. For a large part, it did not become an administrative part of The Netherlands (or Belgium) until well into the 19th century (Kessels-van der Heijde 2002). The following map shows the positioning of Limburg (the grey spotted area) with respect to the other parts of the Netherlands (the western most part of Limburg is Dutch, the eastern part is Flemish):

![Map of the Netherlands showing the positioning of Limburg](image)

The GTR data were mainly used to check the robustness of the Maasbracht intuitions. With this goal in mind, we compared the feminine forms of the adjectives klein ‘small’, oud ‘old’, goed ‘good’, heel ‘very’, rijp ‘ripe’, rond ‘round’, lang ‘long’, scheef ‘oblique’ and hoog ‘high’ with their neuter or citation forms in the database. After filtering out those forms for which the

---

25 Data come from a few different sources. These are, first, Hermans (1994), presenting a wealth of native speaker’s intuitions on one individual dialect, Maasbracht Dutch, plus a very insightful analysis of some of these data, on which we will draw. Secondly, we use the so-called Goeman-Taeldeman-Van Reenen (GTR) database, a large survey on the phonology and morphology of dialects in The Netherlands and Flanders in the 1980s and 1990s. Maasbracht is approximately in the center of this area, as the map in (142) shows.

26 Neuter forms are those forms given in attributive position with a neuter noun; citation forms are those words which were elicited when the adjective was given in isolation, without any noun. We used the neuter forms for ‘klein’, ‘oud’ and ‘geel’ and the citation forms for the
4.2. Tones and adjectival inflection: data

tones were not transcribed, or not transcribed in an understandable way, we obtained 473 neuter-feminine pairs, with the following distribution (HH=level high tone, HL=falling tone):

<table>
<thead>
<tr>
<th>Tone on neuter</th>
<th>Tone on feminine</th>
<th>Number of adjectives</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>HL</td>
<td>157</td>
<td>.33</td>
</tr>
<tr>
<td>HH</td>
<td>HH</td>
<td>64</td>
<td>.14</td>
</tr>
<tr>
<td>HL</td>
<td>HL</td>
<td>246</td>
<td>.52</td>
</tr>
<tr>
<td>HL</td>
<td>HH</td>
<td>6</td>
<td>.01</td>
</tr>
</tbody>
</table>

It is easy to see that the number of falling neuter - level high feminine pairs is extremely small, especially given the fact that more than half of all the adjectives have an ‘underlying’ falling tone in the neuter. Furthermore, if we look at these six cases in more detail, we see that two of them can be discarded out of hand, in one case since the transcriber has noted that he was not sure about the tone, and in another case because a different adjective was used in the neuter than in the feminine. This leaves us with only 4 pieces of data (out of 473) with a falling-level high pattern for which we will not be able to provide a solution.

Further analysis shows that 46 out of the 64 level - level patterns are found for one single adjective, *rijp* ‘ripe’, the only one in our sample which ends underlyingly in a voiceless obstruent.\(^{27}\) This will turn out to be significant in the following section. As a matter of fact, given that we have reliable tonal data on 59 dialects for *rijp*, we can say that for this word level high-level high is the dominant pattern.\(^{28}\)

All in all we can make the following observations:

\[(144)\]

a. If the stem ends in a voiced obstruent, a sonorant, or a vowel we find two patterns:
   i. neuter: falling, feminine: falling
   ii. neuter: level high, feminine: falling

b. If the stem ends in a voiceless obstruent (*rijp* ‘ripe’), we find level-level patterns (possibly next to the other two)

Other adjectives. The reason why we did not make a uniform choice was purely pragmatic: there are not enough pure neuter/feminine pairs in the GTR database. Given the fact that both the neuter form and the citation form reflect the underlying representation, we trust that this choice does not affect the argument.

\(^{27}\)From the orthography, it might appear that *scheef* ‘oblique’ ends in a voiceless fricative, but this voicelessness is not underlying. It is a quirk of Dutch orthography that final devoicing is represented in fricatives, but not in stops. The word *rijp* actually has a fricative in some of the tonal Limburg dialects — *rijf* —, and this behaves as underlyingly voiceless. We will discuss some the implications of final devoicing in section 4.4.

\(^{28}\)Eleven dialects show a level high-falling pattern, and one dialect shows a falling-falling pattern. These will be left out of consideration.
This conforms to the findings of [Hermans, 1994]. As we have already seen above, this author describes a pattern in which underlyingly level high tones turn into falling tones on the surface, while underlyingly falling tones do not change at all. But [Hermans] also notes that “it is a curious fact of Limburgian morphophonology that tonal alternations can never take place when the base ends in a voiceless obstruent.”

[Hermans] gives the following Maasbracht facts by way of illustration:

<table>
<thead>
<tr>
<th></th>
<th>neuter</th>
<th>feminine</th>
<th>masculine</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>říık</td>
<td>říıkə</td>
<td>říıkə</td>
<td>‘rich’</td>
<td></td>
</tr>
<tr>
<td>nááks</td>
<td>nááksə</td>
<td>nááksə</td>
<td>‘naked’</td>
<td></td>
</tr>
<tr>
<td>záát</td>
<td>záátə</td>
<td>záátə</td>
<td>‘lame’</td>
<td></td>
</tr>
</tbody>
</table>

Although it is not true that all Limburg dialects display this ‘curious fact’ — we have just seen there are a few dialects where an alternation was found after all —, it is true for the majority, and we take this to be an absolute fact for Maasbracht. The generalisation was, incidentally, already made by [van Wijk, 1935]. Given that we have sufficient detailed native speaker evidence only for the Maasbracht dialect, we will concentrate on this dialect in what follows; see [Hinskens & Muysken, 1986] for a thorough analysis of a slightly different system.

One fact will turn out to be absolutely crucial for our present purposes: in the examples in (145), a schwa shows up on the feminine suffix. This schwa is crucially lacking in the examples in (139). We thus can make the following generalisation:

(146) a. if the feminine has a level high tone, it also has a schwa
     b. if the feminine has a falling tone, the schwa does not show up, regardless whether there is alternation in the paradigm or not

This is the correlation that will form the core of our discussion in the next two sections.

### 4.3 A representational analysis

We may simplify the representations in (141) in a number of ways. First, if we consider the low tone in the level high pattern as a phonetic effect, or as the effect of the OCP, we may further simplify this pattern into HH, which then contrasts with HL. We have of course already implied this in our discussion above, by introducing the term ‘level’ high tone.

Notice also that both tones feature a high tone on the first mora. It is true that these lexical tones are realized on exactly one syllable in every word: the syllable with main stress. In other words, the initial tone seems to be uniquely
4.3. A representational analysis

due to some principle relating high tone and stress, which of course has been known for a long time in the phonological literature (cf. [Hulst & Smith][1988] for an overview):

(147) **Pitch:** The head mora of the syllable with primary stress needs to have a high tone.

One can view (147) as an Optimality Theoretic constraint (possibly formalized along the lines of [de Lacy][1999][2002], in which case this constraint is inviolable in the grammar of Limburg. It is only the tone on the second mora in the main stressed syllable which can be either H or L, subject to lexical specification.

It is most likely that of these two, the Low tone is the phonologically marked one. For instance, if we have a minimal pair of words, one with a level tone and the other with a falling tone, and if one of those two is a function word and the other one a lexical word, it will be typically the one with the level high tone which is the function word and the one with the falling tone which is the lexical word (e.g. bī ‘with’ - bii ‘bee’, ziī ‘she’ - zii ‘silk’). If we assume that function words are usually phonologically less marked than lexical words, we can understand these patterns as an indication that H will be the default tone.29

The next step in our analysis is that the neuter suffix is a truly empty morpheme with neither a schwa nor a mora nor a tone. The masculine suffix we assume to consist of a schwa plus a low tone. The feminine suffix, on the other hand, would consist of two parts: an empty vocalic position, and a tone.30

(148) Neuter | Feminine | Masculine
--- | --- | ---
σ | σ | σ
\(\hat{\imath}\) | \(\hat{\imath}\) | \(\imath\)
∅ | L | L

**Neuter and masculine suffixes**

Let us first consider the neuter and the masculine suffixes. We can either of these add these to either a stem with an underlying low tone, or to one with

---

29Laura Downing (p.c.) points out that this analysis could be taken to imply that high tones also surface on stressless syllables. Usually they are taken to be toneless in the dialectological literature. In order to explain this, we we will invoke TONESTRESS below, requiring all tones to be in a stressed syllable.

30See [van Oostendorp][2005b] for extensive argumentation for the existence of empty-headed morphemes in dialects of Dutch.
an underlying high tone (or no underlying tone at all). This gives us four possibilities, two for the neuter and two for the masculine:

1. If we add a neuter (empty) suffix to a lexical form with a low tone, the underlying low tone will show up on the second mora. The reason for this is that tones need to be within the main stressed syllable, and the first mora is already occupied by a high tone, according to PITCH:

   (149)   a. TONESTRESS: Tones need to be in the syllable bearing main stress
           b. MAXTONE: Do not delete tones
           c. PITCH ≫ TONESTRESS, MAXTONE


<table>
<thead>
<tr>
<th>/k ál m +Low/ + ∅</th>
<th>PITCH</th>
<th>TONESTRESS</th>
<th>MAXTONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [k ál m]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. [k ál m]</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>c. [k ál m]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. If we add an empty neuter empty suffix to a lexical form with an underlying level high tone, the result is a level high tone. If the adjective does not have any tone at all, we may surmise that the form will also turn up with a level high tone — this is the sense in which this tone is ‘unmarked’. In order to achieve this result, we assume that every mora in the stressed syllable needs to have a (high) tone. If the relevant constraint is ranked below MAXTONE, this does not affect the results we have obtained so far:

(151) STRESSSTOTONE: All moras in the syllable bearing main stress must bear tone.

(152) /l a á m/ + ∅ | PITCH | TONESTRESS | MAXTONE | STRESSSTOTONE
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [l a á m]</td>
<td>*!</td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>b. [l a á m]</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c. [l a á m]</td>
<td>*!</td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>d. [l a á m]</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>e. [l a á m]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The forms in (152a), (152c) and (152d) do not have a tone on one of the moras in the stressed syllable; they are therefore unacceptable. The choice is between (152b) and (152e). The latter wins, because it has high tones on all moras of the stressed syllable. From this we can conclude

---

31We do not have evidence yet for TONESTRESS ≫ MAXTONE, but we will see this below.
that Pitch is a more specific version of StressToTone, which again could be formalized along the lines of de Lacy (1999, 2002).

3. If we add a masculine (low tone) suffix to a lexical form with a low tone, we will get a low toned form. At present, we have no clue as to which of the two underlying low tones is actually surfacing:

(153)

<table>
<thead>
<tr>
<th>/kalm + Low/ + /a+ Low/</th>
<th>Pitch</th>
<th>ToneToStress</th>
<th>MaxTone</th>
<th>StressToTone</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [kəlma]</td>
<td>#!</td>
<td>**</td>
<td>#!</td>
<td>**</td>
</tr>
<tr>
<td>b. [kɪlma]</td>
<td>#!</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. [kɪlma]</td>
<td>#!</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>d. [kʊlma]</td>
<td>#!</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>e. [kʊlma]</td>
<td>#!</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>f. ɪ[kɪlma]</td>
<td>#!</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

Note that this paradigm provides us with information on the ranking of ToneToStress and MaxTone: one of the two underlying tones has to be deleted here, because it cannot surface in a non-stressed position.

4. If we add a masculine (low tone) suffix to a lexical form without a tone, the low tone of the suffix will surface, and a falling tone will ensue:

(154)

<table>
<thead>
<tr>
<th>/laam/ + /a+ Low/</th>
<th>Pitch</th>
<th>ToneToStress</th>
<th>MaxTone</th>
<th>StressToTone</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [laamə]</td>
<td>#!</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>b. [ləamə]</td>
<td>#!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c. [ləamə]</td>
<td>#!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>d. [lʊamə]</td>
<td>#!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>e. [lʊamə]</td>
<td>#!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>f. ɪ[ləamə]</td>
<td>#!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

This concludes our analysis of the masculine and neuter forms (of stems not ending in a voiceless obstruent; we will return to the latter in section 4.3). Notice that the set of constraints that we require is relatively small and furthermore fairly ‘natural’, at least from a typological point of view. The only constraints we need are those establishing a relation between metrically strong positions and tones – and preferring high tones over low tones in this respect.

**Feminine suffixes**

We now turn to the feminine suffix, for which I propose that it consists of an empty mora plus a low tone. Independent phonological constraints will need to interpret the empty vocalic position. We propose that the default choice is that it simply does not get a phonological interpretation at all. In this way, it satisfies better the constraints of the family *Struc, instantiated here as *Schwa. If faithfulness (in particular a constraint against deletion of vowels, Max-V) dominates this markedness constraint, masculine forms will not be affected:
4.3. A representational analysis

Yet feminine suffixes can do without the schwa without being unfaithful (we use /µ+Low/ in the tableaux to represent the feminine suffix of which the ‘real’ structure is the one given in (148)):

(156)  

<table>
<thead>
<tr>
<th>/laam/ + /µ + Low/</th>
<th>MAX-V</th>
<th>*SCHWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [lāːm]</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. [lāːmə]</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

The Low tone in the feminine is underlyingly present, just like in the masculine, and therefore will show up wherever it can. The difference between neuter, masculine and feminine thus is a difference in lexical specification of the respective morphemes. Most important, at present, is the difference between the neuter and the feminine: whereas the latter has an underlying low tone, the former does not.

Without having to stipulate additional constraints we can now derive the pattern for the feminine suffix. If we add it to an adjective with an underlying low tone, one of the two low tones surfaces, and if we add it to an adjective without an underlying tone, the low tone of the suffix surfaces. The empty position will stay empty for faithfulness reasons just outlined. All of this is exactly like what we found for the masculine suffix. The only difference is that in this case we do not find a schwa:

(158)  

<table>
<thead>
<tr>
<th>/laam/ + /µ + Low/</th>
<th>PITCH</th>
<th>TONEToSTRESS</th>
<th>MAXTONE</th>
<th>STRESSToTONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [laam]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. [lāːm]</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. [lāːm]</td>
<td></td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>d. [lāːmə]</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>e. [lāːmə]</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

(155)  

<table>
<thead>
<tr>
<th>/laam/ + /ə+ Low/</th>
<th>MAX-V</th>
<th>*SCHWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [lāːm]</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. [lāːmə]</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
4.3. A representational analysis

Stems in voiceless obstruents

Let us now turn to stems ending in a voiceless obstruent. These forms never alternate: neuter, masculine and feminine forms all have a level high tone. In addition a schwa shows up on the feminine, as the facts of (145), repeated here for convenience, show:

(145)   neuter  feminine  masculine
  rík  ríko  ríkə  ‘rich’
  nááks  nááksə  nááksə  ‘naked’
  záát  zááťə  zááťə  ‘lame’

The fact that low tones are avoided on syllables ending in an underlyingly voiceless obstruent is obviously in need of an independent explanation. There are reasons to assume that some constraint is active in the phonology of Limburg, disallowing the combination of low tone and voicelessness (see [Hermans & van Oostendorp, 2001; Hinskens & van Oostendorp, 2005] for more discussion). One way to formalize this, is by assuming an implicational relation such as the following:

(159)  L ⊃ [+voice]: A Low tone implies a feature value [+voice]

We could read this constraint as one requiring consonants always to be voiced in the vicinity of low tones, or as low tones dispreferring to land next to voiced consonants. This constraint can be seen as phonetically grounded in the sense that there is a clear connection between voicing of consonants and lowering of F0 values ([Maddieson & Hess, 1987] — a more radical version of this analysis would have it that Low and [voice] are the same feature, see [Halle & Stevens, 1971; Bradshaw, 1999; Harris, 1994] among others, for arguments in favour of such a position. This constraint, then, directly blocks low tones from surfacing, if it dominates the faithfulness constraints on tone.32 This is illustrated in the following tableau for the masculine form of rijk ‘rich’ (assuming, irrelevantly, that the adjective itself does not carry a low tone):

(160)  

<table>
<thead>
<tr>
<th>/rijk/ + /a+ Low/</th>
<th>L ⊃ [+voice]</th>
<th>MAXTONE</th>
<th>STRESSTOTONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [rika]</td>
<td>*</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. [rũka]</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. [rika]</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

32 In section 4.2, we noticed that there are a few dialects which do seem to display alternations in this case. If those data turn out to be right, this could be a result of a reranking of the relevant constraints.
The only form which can win has a high tone on the second mora. A low tone is disallowed next to a voiceless obstruent, and the second mora needs some tone because it is in a stressed position. Therefore the masculine suffix is realized only partly in this particular position.

How about the feminine morpheme? If things would work the same way as for the masculine forms, we would select *[ríık]*, which would be homophonous to the neuter form. This apparently does not happen. Notice, however, that the homophony itself is not always a fatal problem, since the feminine and neuter forms of calm with a falling tone are also homophonous: *[kólın]*. Furthermore, the feminine form which is selected, *[rííka]* is homophonous to the masculine. So avoidance of homonymy within the adjectival paradigm cannot serve as an explanation.

In order to understand what is going on, I propose to refer to the REALIZE-MORPHEME, in accordance with a tradition in the phonological literature and define it as a special type of faithfulness constraint.

(161) REALIZE-MORPHEME (RM): For every morpheme in the input, some phonological element should be present in the output.

This constraint could be interpreted in the light of recent work on Optimality Theory in semantics and pragmatics; see for instance Buchwald et al. (2002) and the contributions to Blutner & Zeevat (2004). A central notion is recoverability (there is some debate in the literature on the correct terminology and the proper way of implementing this idea). This notion explains, for instance the ‘reduction’ of nominals to pronominals. If somebody says ‘He is coming’ in stead of ‘John is coming’, she may be satisfying the requirements of *STRUC, since pronouns contain less information (hence less structure) than nouns (or proper names). Why don’t people then reduce all nouns all the time? The answer is recoverability: a higher ranking constraint demands that we can only use ‘he’ if from the context we can recover the extra information that we are talking about John.

I propose that we have something very similar here in the phonology-morphology interface. We usually prefer the schwaless form for the feminine, since it contains less structure. However, in the case of stems ending in voiceless obstruents, this would mean that the suffix is not realized at all (it contains only a Low tone, and this cannot surface). But that would mean that the morpheme is completely unrecoverable: there is no trace in the phonological surface form which shows that it is there. In this case, then, we choose the allomorph with schwa, which will still be recoverable.

Let us now see how this constraint affects the analyses of the neuter, masculine and feminine forms for words ending in a voiceless obstruent. For
neuter forms, evaluation of RM is vacuous: since there is no underlying material at all, nothing can serve as a representative of the neuter suffix on the surface. For the masculine, there are in principle two elements which can satisfy RM and since the schwa always surfaces for independent reasons, the tone is not necessary, so that also in this case addition of the constraint does not affect the argumentation.

For the feminine form, we now have to assume that recoverability outranks structural markedness, i.e. $RM \gg^* SCHWA$:

\begin{align*}
\text{/riik/} & + /\mu + \text{Low}/ & \text{RM} & \text{*SCHWA} \\
\text{a. [r`i`ık]} & & \ast' & \\
\text{b. } & & & 
\end{align*}

One way of picturing the working of RM is by assigning a subscript to the elements of every morpheme. RM then has it that every subscript has to be present on some element on the surface:

<table>
<thead>
<tr>
<th>underlying representation</th>
<th>bad surface form</th>
<th>good surface form</th>
<th>good surface form</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_i a_i l_i m_i a_j$</td>
<td>$r_i a_i l_i k_i$</td>
<td>$r_i a_i l_i k_i$</td>
<td>$r_i a_i l_i k_i$</td>
</tr>
<tr>
<td>$L_j$</td>
<td>$L_j$</td>
<td>$L_j$</td>
<td>$L_j$</td>
</tr>
</tbody>
</table>

The underlying representation has two morphemes, corresponding to two subscripts, $i$ and $j$. The candidate surface form in the middle is bad because it has only one of those two subscripts. The two forms at the right hand side obey RM, because they have both subscripts. (The rightmost one will eventually be chosen because of the phonological constraint $L \supset [+\text{voice}]$.)

Note that the required visibility, if seen this way, is somewhat abstract, because it is intermediated by subscripts. This provides us with a way to distinguish between the two possible output representations for forms such as the feminine for calm:

<table>
<thead>
<tr>
<th>underlying representation</th>
<th>bad surface form</th>
<th>good surface form</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_i a_i l_i m_i$</td>
<td>$k_i a_i l_i$</td>
<td>$k_i a_i l_i$</td>
</tr>
<tr>
<td>$L_i$</td>
<td>$L_i$</td>
<td>$L_i$</td>
</tr>
</tbody>
</table>

Even though the two potential output forms are homophonous, we now have a theory-internal reason to choose for the rightmost one: this one still contains all indices of the underlying representation. For this reason, there is no need to insert a schwa (or preserve it) in this case. Note that the ‘bad’ surface form will surface in the derivation of the neuter, simply because the rightmost form will be unavailable here.\footnote{This means either that we restrict the Generator function in such a way that it cannot add morphological affiliations to segments (this was called Consistency of Exponence in Prince Quantum Grammar).}
This completes our analysis of tone in adjectival inflection in Limburg Dutch. We repeat the constraint rankings we have called upon in (165):

(165) a. $\text{Pitch} \gg \text{ToneToStress} \gg \text{MaxTone} \gg \text{StressToTone}$

b. $L \supset [+\text{voice}] \gg \text{MaxTone}$

c. $\text{Max-V,RM} \gg *\text{Schwa}$

The subhierarchies in (165a) and (165b) regulate the distribution of tone, and (165c) regulates the occurrence of schwa. The two processes are almost independent, except that deletion of underlying tone will affect RM in exactly one case: that of feminine suffixes before voiceless obstruents.

### 4.4 Paradigms and representations

#### Lenition and final devoicing

The previous section presented the main line of analysis. In this section we will fill in a few details, and compare our analysis to two alternatives.

In addition to the tonal behaviour already mentioned, the feminine form of the Limburg Dutch adjective is different from the neuter in another respect: stem-final underlying /d/ lenites to [j]:

(166)

<table>
<thead>
<tr>
<th>neuter</th>
<th>feminine</th>
<th>masculine</th>
</tr>
</thead>
<tbody>
<tr>
<td>rò̆d</td>
<td>rò̆ɔj</td>
<td>rò̆ɔj@</td>
</tr>
<tr>
<td>rì̆̆k</td>
<td>rì̆̆kɔ</td>
<td>rì̆̆kɔ@</td>
</tr>
<tr>
<td>zò̆̆t</td>
<td>zò̆̆tɔ</td>
<td>zò̆̆tɔ@</td>
</tr>
</tbody>
</table>

`red`

Lenition of this type usually only happens to /d/’s in (intervocalic) onset position in dialects of Dutch (such as in the masculine form here Zonneveld, 1978). Yet in the feminine form chosen here, there is no vowel.36

Notice that this fact gets a natural explanation under the analysis proposed here. Since the feminine suffix contains an empty vocalic position, the /d/ will still be literally intervocalic in the feminine, even if one of the two vowels is not pronounced, and hence be prone to lenition. The neuter does not provide such a position, on the other hand, and therefore the /d/ at the end of the neuter is not subject to lenition.

There is also an alternative analysis, for which we first have to consider the most likely output candidate for roõ ̆d ‘red’ without lenition. Hitherto we have assumed that this is the following:

& Smolensky (1993), or that the faithfulness constraint RM only looks at those subscripts which are already present underlyingly: none, in the case of the neuter.

36Furthermore, this lenition is pervasive in the Limburg dialect area. The GTR database contains 58 Limburg Dutch dialects with reliable data on the adjective goed ‘good’. None of these end in a plosive (whereas all the neuter forms do). For 24 dialects, the final segment is transcribed as [i]; 28 dialects have [j], and the rest have [w], [u] or [y].
However there is something definitely uncomfortable about this analysis and this is that Limburg Dutch, like all Dutch dialects has a process of final devoicing, which is to say that the final segment is not [d], but rather devoiced [d] (or [t]). The problem with this obviously is that we have evidence that (underlyingly) voiceless obstruents such as /t/ do not permit Low tones in front of them.

This implies that we have to distinguish between underlyingly voiceless and devoiced consonants. One way to achieve this effect is by following Ernestus (2000) and assume that while voiced consonants are [+voice] and voiceless consonants are [-voice], the result of final devoicing has no specification for voicing at all (Ernestus, 2000 gives a range of phonetic, phonological and experimental evidence for this). In that case, we could split up the constraint called $L \supset [+voice]$ above into two parts:

(168) a. $^* L^{-voice}$: disallowing the combination of [-voice] with a low tone

b. $L \supset [+voice]$: requiring low tones to be accompanied by [+voice] segments.

An underlyingly voiceless [t] would violate both constraints, whereas a devoiced [d] would only violate the second one. If we then put the constraint in (168a) at the inviolable position we have awarded to the voicing constraint in the previous section, and demote $L \supset [+voice]$ to a much lower position, the result is that devoiced consonants are more permissive, and will usually tolerate low vowels before them.

Yet among these devoiced consonants, [d] happens to be the only one which can avoid violating the second constraint at a relatively low cost, viz. by turning into a sonorant [j]. In this way, then, the lenition can be seen as an Emergence of the Unmarked effect on the constraint in (168b): falling tones are permitted before devoiced consonants, but only in case nothing can be done to change those devoiced consonants into something more acceptable (see Hinskens & van Oostendorp 2005 for an elaboration on this idea).

Paradigms or morphemes

Having now set up a representational OT analysis, we may compare it to another OT account of the same phenomenon, one in terms of paradigms, proposed by Alderete (1999).
Alderete (1999) gives a purely morphological approach based on output-output correspondence relations. To be more precise, Alderete (1999) defends a notion of Anti-faithfulness: some morphological forms – e.g. forms in a paradigm – desire to be different from other surface form in some properly described way. (Again, we will not go into all of the technicalities of the approach.)

For the Limburg data, Alderete assumes that low tones are absent altogether: a falling tone is represented with a high tone on the first mora, and nothing on the second mora. This makes them thus less marked than level high tones. Further, there is a constraint ¬NO-FLOP-TONE, which informally states the following:  

\[(169) \quad \neg \text{NO-FLOP-TONE} \text{ If a segment } s_1 \text{ is linked to a tone } T_1 \text{ in the neuter, a corresponding segment } s_2 \text{ should not be linked to a corresponding tone } T_2 \text{ in the feminine (and masculine)}\]

The following gives the input-output pair for the masculine form of *lame* as well as the neuter form (which does not change from input to output):

\[(170) \quad \begin{array}{c|c|c}
\text{input} & \text{output} & \text{neuter} \\
1 \text{ a a m+ } & 1 \text{ a a m+ } & 1 \text{ a a m} \\
\downarrow H & \downarrow L & \downarrow H \\
\end{array}\]

The output form of the masculine has changed from input to output, because ¬NO-FLOP-TONE requires the tonal association of the masculine to be different from that of the neuter. The reason why this affects the last mora of the word is because of tonal alignment: within syllables, tones prefer to be at the left edge. (We interpreted the same facts to mean that the masculine suffix has a low tone, which the neuter suffix does not have.)

¬NO-FLOP-TONE does not take effect if the neuter has a falling tone. In that case there is only one tonal association: of the high tone to the first mora, but this cannot be undone because of a high-ranking constraint (we argued this to be PITCH, for Alderete (1999), it is again left alignment of tone in syllables).

Alderete (1999) mentions the following advantages of his approach (p. 226):

\[(1986), \text{ but we will not discuss this here because it deals with a dialect with a slightly different pattern, and favours an analysis which is based on theoretical assumptions very different from the one presented here.}\]

\[\text{Alderete (1999)’s approach is based on antifaithfulness of the feminine form with respect to the neuter or citation form. Note that it would also be possible to construct a paradigm uniformity approach with faithfulness to the masculine form. As far as I can see, this would have the same properties as the Alderete (1999)’s theory – it would share its advantages, but also its problems.}\]
1. “The analysis presented here accounts for accent purely in terms of H tones, and as emphasized above, the analysis is in line with recent approaches to tonal accent systems like the one given in Pulleyblank 1986 for Tonga.”

2. “A second point in favor of [Alderete (1999)]’s analysis is that it relates a wide range of morphologically triggered shifts as effects of a specific type of Anti-Faithfulness. Thus, the loss of a link in the dragging tone mutation is treated on a par with the obligatory shifts found in Japanese and Aguaruna.”

3. “A final argument in favor of the account of the accentual mutation in terms of AntiFaithfulness is that it explains the relation between the properties of the accent shift with independently needed constraints. ”

I believe that none of these arguments hold. Ad 1, it can be observed that there are independent reasons to assume that Low tone is the marked tone in Limburg Dutch, and the morphologically active one. It serves to mark the plural (which thus has a falling tone) from the singular (with a level tone) in nouns, for instance (e.g. bèın - bèın ’leg - legs’), and in all minimal pairs where one of the two forms is a lexical word and the other one a function word, the lexical word has the falling tone and the function word the level tone (e.g. zǐí - zǐí ’she - silk’). This can be understood if function words are supposed to have an unmarked phonological structure, whereas lexical words are more marked, and if the Low tone is marked, i.e. present in the phonological representation. Whatever the merits of the assumption that only High tones are present for the analysis of Tonga, it seems to be jumping to conclusions to assume that this should carry over to all other languages, including Limburg dialects.

Points 2 states that the anti-faithfulness analysis has as an advantage that it relates the Limburg facts to those of the morphologies of other languages. But the same seems to be true for an analysis which holds that neuter and feminine have different suffixes.

Finally, Alderete (1999) mentions as an advantage of his analysis that it uses phonological constraints which are motivated independently; but the same is true for the analysis presented here. In sum, none of the ‘favourable’ properties mentioned by Alderete (1999) seem to be convincing enough to blow out the approach mentioned here.

In return, [Alderete (1999)] does not discuss the interaction with voicelessness on obstruents, and it is hard to see how those facts could be incorporated into a paradigmatic approach. We might be able to constrain ¬NO-FLOP-TONE in such a way that it does not affect words ending in a voiceless obstruent, but even then, there is no reason why a schwa should appear at the same time. Allomorphy is not a notion to which we can refer, since this approach does not refer to morphemes at all: the tonal shift if encoded in the morpheme-specific constraint ¬NO-FLOP-TONE, not in the representation of
any kind of constraint.

On a formal level, we argue that interparadigmatic faithfulness is too abstract and too powerful a formal device to incorporate into our theory too lightly. The approach defended here might be slightly abstract since it involves an empty vocalic position. At the same time, the antifaitfulness approach is abstract in many more ways. Not only does it posit ‘toneless’ mora’s in stressed syllables, which then have to be interpreted as low, but also do we have to assume correspondence relations among individual segments and tones in words - and none of these can be observed phonetically anymore than morphological superscripts can.

Since the latter approach is more parsimonious, and at the same time seems more succesful from an empirical point of view, we conclude that it is preferable over its current competitors. We claim that it is the interaction between phonology and morphology which gives us exactly the pattern we find in Limburg Dutch adjectival inflection.

5 Derived environment effects and colouring

5.1 Some examples of derived environment effects

One important aspect of the phonology-morphology interface which has been discovered in the second half of the twentieth century, is that some phonological processes only apply in derived environments, that is to say they do not apply to underlying monomorphemic forms (Kiparsky, 1973, 1993; Kenstowicz & Kisseberth, 1977; Anttila, 2005). Traditionally, two types of derived environment are recognized: phonologically and morphologically derived environments. Morphologically derived environments consist of material from more than one morpheme; phonologically derived environments can be monomorphemic, but at least one of the elements has to be created by an earlier phonological rule or process. In as far as these effects are real, phonological theory needs to account for them, and I propose that Coloured Containment is optimally suitable for that.

In order to illustrate this point, I will consider five rather well-known instances of Derived Environment Effects (DEE), from Korean, Turkish, Finnish, Polish and Tuscan Italian.

Korean has a rule of palatalisation, affecting coronal stops before front high vowels, roughly formulated as in (171a) and illustrated in (171b). This rule only applies across morpheme boundaries however; it does not apply in the form in (171c), where the /t/ and /i/ are already adjacent at the underlying level (Iverson, 1993; Polgárdi, 1998; Rhee, 2002, to mention just a few recent sources).

(171) a. t, tʰ → ŭ, ŭʰ / _ i
5.1. Some examples of derived environment effects

b. hæ tot+i → hæ doq ‘sunrise-NOM’
c. mati → madi ‘knot’

The second example comes from Vowel Harmony in Turkish. As is widely known, Turkish has both backness and roundness harmony (subject to different phonological constraints) (172a). Roots can be disharmonic; although the forms in (172b) are of foreign etymology, they have integrated into the Turkish phonology in other respects. (172c) shows that epenthetic vowels are

\[
\begin{array}{c|c|c|c|c}
\text{nom.sg.} & \text{gen.sg.} & \text{nom.pl.} & \text{gen.pl.} \\
\hline
\text{‘rope’} & \text{ip} & \text{ipin} & \text{ipler} & \text{iplerin} \\
\text{‘girl’} & \text{kız} & \text{kızın} & \text{kızlar} & \text{kızların} \\
\text{‘face’} & \text{yüz} & \text{yüzün} & \text{yüzler} & \text{yüzlerin} \\
\hline
\text{vali} & \text{‘governor’}, \text{kitap} & \text{‘book’}, \text{hareket} & \text{‘movement’}, \text{hesap} & \text{‘bank account’}, \text{bobin} & \text{‘spool’}
\end{array}
\]

c. careful form colloquial form
‘fetters’ pranga pranga
‘cruiser’ kruvazör kruvazör

We can see this blocking of productive harmony to roots as a DEE; the fact that a vowel within a root can be subject to harmony only if it is epenthetic lends further credence to such a view.

A third example of a DEE comes from Colloquial Helsinki Finnish Vowel Coalescence, as has been recently demonstrated by Anttila (2005). This particular dialect of Finnish has a rule of Vowel Coalescence, which turns /makea/ ‘sweet’ for instance into [má.keː]; the rule is optional, so that [má.keːa] is also possible. Anttila (2005) notes that Vowel Coalescence is subject to a quantitative DEE effect:

- Vowel Coalescence is categorically blocked in non-derived environments if the structural change is highly marked. If the structural change is unmarked, it may apply.
- Vowel Coalescence is quantitatively dispreferred in derived environments if the structural change is highly marked. If the structural change is unmarked, it is quantitatively preferred.

All in all, we will thus find more vowel coalescence in derived environments. It is quite interesting that there is quantitative variation of this type, and Anttila (2005) offers an analysis of the statistic effect, which falls beyond the scope of our present concerns. But the fact that there is variation of this quantitative type at all, shows that the DEE is real within the synchronic grammar.

A fourth example of a DEE is Polish Spirantisation (Rubach, 1984); as a matter of fact this is at present probably the most widely discussed instance of this phenomenon within Optimality Theory because of the work
5.2 Colours show Derived Environments

The process is slightly more complicated: an underlying /g/ turns into a [ʒ] before an [e]. This can be seen as palatalisation of /g/ to [j], and subsequent spirantisation to [ʒ] (173a). It is this spirantisation which is subject to a DEE, since underlying /j/ does not undergo it (173b).

(173)  
   a. /rog+ek/ → [rożek], *[rojek]  
   b. /briż+ek/ → [brijek]

Our last example comes from the Tuscan dialect of Italian. Krämer (2005) shows that in this dialect, there is a phonemic contrast between /s/ and /z/. This contrast is lifted, however, if the coronal fricative occurs at the edge of the morpheme, in which case the contrast is neutralised to [z]:

(174)  
   a. fu[z]o ‘melted’ - fu[s]o ‘spindle’  
   b. ro[z]a ‘rose’ - ro[s]a ‘itch’  
   c. di[z]onesto ‘dishonest’, di[z]uguale ‘unequal’

What this means, according to Krämer (2005), is that the process of intervocalic s voicing is subject to a DEE: it applies in a morphologically derived environment, but not elsewhere.

5.2 Colours show Derived Environments

Now that we have set up a catalogue of — hopefully representative — examples, we will build an analysis of DEE in constraint-based theory. In order to do this, we need to be able to evaluate both input and outputs; we thus need aspects of both faithfulness and markedness. I argue that the advantage of Coloured Containment is that it offers both in one representation, while in Correspondence, on the other hand, the separation between F and M is too large, causing problems of locality. From the rule-based work on DEE, we know that there are several diagnostics for DEE-sensitivity (lexical rules, structure-building, etc) but also that there are problems with these diagnostics. I propose a different diagnostic here: DEE will always involve spreading. This seems to be tenable at least for the examples studied here.

The application of Coloured Containment to DEE can be easily illustrated on the basis of Turkish. Consider the following form, the genitive of the word kruvazör ‘cruiser’, with an epenthetic vowel inserted within the first cluster, and a suffix -ın added at the end:

(175)  

\[
\begin{array}{cccccccc}
  & k & a & 1 & r & a & u & a & v & a & a & z & a & ķ & a & r & a & l & b & n & b \\
\end{array}
\]

\[
\begin{array}{cc}
  \text{[round]}_a & \text{[round]}_a
\end{array}
\]
5.2. Colours show Derived Environments

Now let us compare those structures which are allowed, to those which are not:

(176)  

<table>
<thead>
<tr>
<th>Allowed</th>
<th>Not allowed</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_α u r_a u_u a_z a \hat{a}_a r_a r_b n_b )</td>
<td>( k_α u r_a u_u a_z a \hat{a}_a r_a r_b n_b )</td>
</tr>
</tbody>
</table>

A comparison of these forms teaches us that it is possible to associate a feature and a segment of different colour (either because the segment is epenthetic or because the feature is in a different morpheme than the segment), but it is not possible to associate a feature and a segment of the same colour. On the other hand, we obviously do not want our constraint to disallow underlying associations between segments and features, such as that between the two /u/ segments and their feature [labial].

One way of implementing this would be to assume that association lines, like other elements of a phonological representation have a morphological colour. If an association line is underlying, it has the colour of the morpheme to which it belongs; if it is not underlying, it does not have a colour. Reviathioudou (this volume) provides a better formalisation of this, but for our purposes the following constraint will suffice:

(177)  

\textit{Alternation:} If an association line links two elements of colour \( \alpha \), the line should also have colour \( \alpha \).

The structure in (176b) violates this constraint since it links features [round] to a segmental node which is in the same morpheme, hence has the same colour. The structure in (176a) on the other hand is fine, since the features and the segments to which they are linked have the same colour. The underlying association lines in both structures have the same colour as the elements they link (by definition) and hence they are fine as well.

This analysis can be easily extended to the other languages we have discussed. Compare for instance the licit form of palatalisation in Korean on the lefthand side and the illicit form on the righthand side:

(178)  

\( t_c o_a \hat{a}_a i_b \)  
| \textit{[high]}_a | \textit{[high]}_c |

The form on the righthand side is characterised by a new — i.e. colourless — association line linking two elements of the same colour, which is not allowed.
by ALTERNATION. In the form on the lefthand side, on the other hand, the
two elements which are linked have different colours, hence the association
line is allowed.

Similarly, it becomes quite clear what happens in Finnish once we draw
the right association lines:

\[
\begin{array}{cccccc}
  l_a & a_d & s_a & i_a & a_b & r_c a_c & s_c & i_c & a_c \\
  l_a & & & & & & & & \\
\end{array}
\]

(179)

Again, the preferred structure is characterized by new association lines link-
ing elements of different colours, whereas the dispreferred (or disallowed
structure) has a new link between material which is already underlyingly
present in the same morpheme.

In this approach, there thus is no qualitative difference between phono-
logically and morphologically derived environments. In both cases there is
a difference in colour; in phonologically derived environments since one of
the elements is epenthetic, i.e. colourless, and in morphologically derived
environments since different morphemes each contribute their own colour.

The situation is a little more complicated for the case of Polish spiranti-
sation. In the examples just discussed, a whole segment was ‘new’, either
because if was phonologically derived or because it belonged to a different
morpheme. In Polish, this is not the case: the difference between \[\text{roZek}\],
\*[\text{ro\'zek}] and \[\text{br1\'zek}\] is not that the former contains a new segment in any
sense in which the latter does not. The difference is that the former contains
a segment which has been changed.

In order to describe this properly, then, we need to resort to Feature Ge-
ometry and assume that ALTERNATE works at a subsegmental level in this
language.

An independent advantage of introducing Feature Geometry, may be that
it allows us to express the intimate connection between [continuant] and
Place. Within other approaches to the Polish facts (such as [Lubowicz] 2002
2005b) it is essentially a coincidence that the derived environment for spi-
rantisation — a change in [continuant] — is created by palatalisation — a
change in Place feature specification. However, it is well known that there is
an intimate connection between these features, which has been expressed in
various ways; here we will choose the analysis of Padgett(1991), according
which [continuant] depends on the Place features (so it can be a daughter of
Coronal, Velar or Labial).

A few more representational assumptions are in order. In line with a very
long-standing tradition, we assume that palatalisation is ‘assimilation’ with
the vowel on the righthand side. With Rubach(1984) we also assume that the
relevant context for spirantization is the lefthand side, since it happens after vowels and sonorants, but not after obstruents:

\[ (180) \quad \text{r}^\text{og} \ 'horn' \to \text{ro}[3]\text{ek} \ (\text{DIM}), \text{pot}^\text{ega} \ 'power' \to \text{pot}[3]\text{ny} \ (\text{ADJ}), \text{skarga} \ 'complaint' \to \text{skar}[3]y\text{'c} \ 'complain' \]

I will assume that this is progressive assimilation of continuancy (cf. Mas-\[caro, 1983 \] Harris, 1984)

To fully understand what happens in Polish, it is useful to first study the structure of these two words when only palatalisation has applied:

\[ (181) \]

\[
\begin{array}{cccccccc}
\text{R} & \text{O} & \text{A} & \varepsilon & \text{K} \\
\text{Place}_{\text{a}} & & & & & & & \\
\text{Coronal}_{\text{b}} & & & & & & & \\
\text{b} & \text{r} & \text{e} & \text{c} & \text{e} & \text{k} & \text{b} & \text{b} \\
\text{Place}_{\text{c}} & & & & & & & \\
\text{Coronal}_{\text{c}} & & & & & & &
\end{array}
\]

In the case of \( /\text{ro}\v{g}+\text{ek}/ \), the Coronal feature will spread from the \( /\text{e}/ \) to the \( /\text{g}/ \), perhaps because of a constraint which disallows a configuration of a non-palatalized velar consonant followed by a front vowel. This spreading conforms to the constraint \text{ALTERNATION}. For \( /\text{br}\v{j}\text{ek}/ \), no change is necessary, since the structure already conforms all well-formedness constraints. Now we extend the structure and look at the interaction of this with spirantisation.

\[ (182) \]

\[
\begin{array}{cccccccc}
\text{R} & \text{O} & \text{A} & \varepsilon & \text{K} \\
\text{Place}_{\text{a}} & & & & & & & \\
\text{Coronal}_{\text{b}} & & & & & & & \\
\text{b} & \text{r} & \text{e} & \text{c} & \text{e} & \text{k} & \text{b} & \text{b} \\
\text{Place}_{\text{c}} & & & & & & & \\
\text{Coronal}_{\text{c}} & & & & & & &
\end{array}
\]

\[ \text{Spirantisation is allowed in } [\text{ro}\v{g}\text{ek}], \text{since the feature } [\text{continuant}] \text{ links to a feature } [\text{Coronal}] \text{ which has a different colour. Spirantisation is not allowed in } [\text{ro}\v{j}\text{ek}], \text{on the other hand, since the place feature and } [\text{continuant}] \text{ have the same colour in this case.} \]

Given all this, the analysis of the DEE in Tuscan has become trivial. If we can see spirantisation as the result of feature spreading, we can certainly also see intervocalic voicing as such. The voicing feature will however only spread to new positions — such as those created by prefixation — but not to old positions — such as those internal to the stem.\[39\]

We have thus seen that all cases of DEE from the previous section are the result of implementing a feature spreading analysis in a Coloured Containment frame. It remains to be seen whether this is true for all DEE, and if not,
how other cases are to be analysed. This also requires a theory of faithfulness to association lines; see Revithiadou (this volume).

5.3 Comparison to other models

Hitherto, various analyses have been proposed to deal with DEE within Optimality Theory, usually based on Correspondence models of faithfulness. We will briefly discuss these alternatives, in order to compare them with the analysis just proposed. By necessity, all these models share some mixing of faithfulness and markedness: we are talking about derived environments, hence phonological environments which are derived (faithfulness) and require some phonological process to apply (markedness).

The most well-known analysis of DEE within OT is couched in terms of Constraint Conjunction (Łubowicz 2000). Under this approach, DEE are seen as the result of conjunction of Faithfulness and Markedness constraints. For the Polish case we have discussed, this is $[^*\xi&\text{IDENT-[CORONAL]}]_{\text{segment}}$, i.e. a markedness constraint requiring spirantisation which is coupled with a faithfulness constraint on palatalisation: if this constraint is high-ranked, only those instances of [\xi] which are the result of palatalisation (hence, which invoke a violation of \text{IDENT-[coronal]}) will be subjected to this constraint.

Several objections can be raised against this account. We have already pointed out that the linking of spirantisation and palatalisation is purely accidental for this account, since any two constraints could be linked. In as far as these processes are indeed intimately linked, the explanation should come from outside the theory of constraint conjunction — possibly from some form of phonetic grounding, although it is not clear how grounding could interact with a purely formal operation such as constraint conjunction. In the Coloured Containment model, only features which are linked in the Feature Geometry can establish a relation of this type.

Further, Constraint Conjunction did not have a theory of locality its original formulation. There is no principled reason why it refers to a violation of \*\xi and of \text{IDENT-[coronal]} within the same segment. This unrestrictiveness predicted unattested long-distance effects. For instance, if we assume that $[^*\xi&\text{IDENT-[CORONAL]}]_{\text{word}}$ would be conjoined within the domain of the word in some language, we would get the effect that palatalisation of a consonant at the end of the word could cause spirantisation of a consonant at the beginning of the word. We this could get derivations of the following type:

$\xiem+ik+ek \rightarrow \breve{z}\!em\!\check{c}\!ek$

These derivations seem unattested, however, and we therefore want to rule them out. A proposal to this effect is provided by Łubowicz (2005b), using ?’s idea of a so-called locus function. The details of this are not important, but
essentially this minimizing the domain of all constraints to the segment: only segments can violate constraints. This restriction seems to general: it is certainly incompatible with the assumption of autosegmental representations. In the other hand, within Coloured Containment, the locality comes for free: palatalisation of one segment in a word will have no representational effect on another palatal in the same word, so a derivation as in (183) can never be derived.

A second approach to DEE in the existing literature, is Comparative Markedness (CM, ?). Like Coloured Containment, CM refers to differences between ‘old’ and ‘new’ material, but it is couched in a Correspondence framework. The central idea of CM is that markedness constraints can refer to ‘old’ violations — those already underlyingly present — vs. ‘new’ violations — those created by Gen. To be precise, for every traditional markedness constraints, CM posits two versions: one referring only to old violations, and one only to new violations. In the case of Polish, this would involve for instance a high ranking constraint *\[\#\]_{\text{New}}^{\text{Old}}: old /\[\#\]/ is unaffected by this constraint, but those instances of [\[\#\]], for instance created by palatalisation, are ruled out by it.

CM is in a certain sense a Correspondence sister (or historically and genetically, a mother) of Coloured Containment theory. Like Constraint Conjunction, CM is however crucially based on the Locus function, which, as we have seen, is incompatible with autosegmental representations. Opinions may of course differ as to whether this is a positive or a negative aspect of the theory, but in the absence of an overwhelming amount of evidence in favour of autosegmental representations, we consider it wise to be conservative.

Furthermore, in order to describe cases of morphologically DEE, CM has to refer to a very special version of ‘newness’, viz. Output-Output_{\text{New}}, satisfied if the violation was not yet there in the underived word. However, introduction of such a constraint type implies simultaneous introduction of a constraint referring to Output-Output_{\text{Old}}, i.e. violations which only count if there is a base form in which they are already present. This has the undesirable result of deriving anticyclicity. For instance, we predict a language like Turkish*, which is like Turkish, except that we find spreading of [-back] only in embedded domains:

(184)  a. Underived form for /kitap/ ‘book’: kitap
        b. Derived form: kitep-lar (SPREAD_{\text{Old}} applies to stem element, which has a base, but not to suffix, which is new).
        c. Derived form: kitep-lar-an (SPREAD_{\text{Old}} applies to kitepler, which has a base, but not to the whole word)

Patterns like this have not been attested, and this seems an undesirable result of CM.

A third approach, Root Faithfulness is proposed as a device to describe Derived Environment Effects in Anttila (2005). This gives very similar effects as
Coloured Containment; for instance, both suffix vowels and epenthetic vowels in Turkish are not part of the root. If we assume FAITH-Root $\gg$ SPREAD, this is why they can both assimilate, whereas vowels in the root are blocked from doing so.

It comes as no surprise that the two approaches yield similar results; we have seen in section ?? that Root Faithfulness constraints need an idea of morphological colour for their implementation. Inversely, Coloured Containment can be seen as an implementation of Root Faithfulness. It is not clear, however, how we could derive the phonological derived environment effects of Polish in a Root Faithfulness approach, without taking recourse to more sophisticated phonological representations, since $[\pm$continuant] and [Coronal] represent different dimensions in an autosegmental approach.

6 Incomplete Final Devoicing

The phonological process of final devoicing is shared by many languages. The following example presents a few well-known cases:

(185) a. Catalan:
   i. gris ‘grey (M)’ - griz ‘grey (F)’
   ii. gos ‘dog (M)’ - gosò ‘dog (F)’

b. Dutch:
   i. kwaa[t] ‘angry (PRED.)’ - kwado ‘angry (ATT)’
   ii. laat ‘late (PRED.)’ - lato ‘late (ATT)’

c. German:
   i. blin[t] ‘blind (PRED.)’ - blindo ‘angry (ATT)’
   ii. bunt ‘colourful (PRED.)’ - bunta ‘colourful (ATT)’

d. Polish:
   i. klup ‘club’ - klubi ‘clubs’
   ii. trup ‘corpse’ - trupi ‘corpses’

e. Russian:
   i. knik ‘book (NOM.SG.)’ - kniga ‘book (GEN.SG.)’
   ii. sok ‘juice (NOM.SG.)’ - soka ‘juice (GEN.SG.)’

The examples in (i) give an alternating pair: a voiceless obstruent at the end of a syllable (in all these cases actually at the end of the word) shows up as voiced when it occurs with an inflectional suffix starting with a vowel — in other words, if it appears in an onset. The examples in (ii) show that this voicing alternation cannot be due to intervocalic voicing, since there are other words ending in voiceless obstruents which do not show a similar alternation.
Interestingly, for all of the languages in (185), it has been claimed in the literature that final-devoicing is phonetically incomplete: the neutralisation is not completely lost, but recoverable in fine-grained phonetic detail. This incomplete neutralisation is cited by Port & Leary (2005) as evidence ‘against formal phonology’. They claim (p.948-949):

The key fact […] is that these word pairs lack an essential property of any symbol token […]: they are neither discretely different nor are they the same. […] It is interesting and probably important that all of the cases of incomplete neutralization mentioned are context-sensitive. That is, they do not represent a general collapse of a distinction in the lexicon. In certain contexts, a distinction has been largely lost, but whatever ‘process’ achieves the neutralization does not completely wipe out the effects of an ‘underlying spelling’ of the lexical items. These cases present troublesome violations of the claim that the phonetics of languages is based on a discrete or digital inventory, and that these discrete phonetic units function like the tokens of a formal system. Instead, speakers can occasionally leave a distinction only partially neutralized by using fine articulatory control.

This assumed lack of success of formalist theories to account for the ‘in-between’ status of neutralised segments, in combination with a few other arguments, leads Port & Leary (2005) to discard any attempt to get to a theory of phonology which is based on a finite alphabet of formal symbols. I argue in this article that incomplete neutralisation does not constitute an argument against formalist theories of phonology. It may constitute an argument against theories of the type proposed in Chomsky & Halle (1968) to which Port & Leary (2005) refer, but this is due to the fact that such theories are based on a view of phonological representations that is too simplistic. I argue that a theory of final devoicing such as proposed in Kooij & van Oostendorp (2003); van Oostendorp (in press), combined with a reasonable Optimality-Theoretic theory of input-output relations as proposed in Prince & Smolensky (1993) and elaborated in van Oostendorp (to appear); Revithiadou (to appear) is sufficiently strong to deal with these facts without any extra ground-breaking assumptions.

6.1 The puzzle: Incomplete final devoicing

It is fair to say that by now there is a fairly substantive literature on the phonetics of final devoicing, and that a large proportion of this literature has shown that this devoicing is incomplete. Even if native speakers will claim that they hear a voiceless obstruent, or will judge that a word ending in a devoiced obstruent rhymes with a word ending in a voiceless obstruent, there
are still two types of indications that the devoicing has not been absolute. The first is that subtle phonetic measurements may still reveal a systematic difference; the second is that in certain types of psycholinguistic tests, native speakers are still able to show a sensitivity to these phonetic subtleties. For instance, if asked to randomly guess whether a given instance \textit{rat} corresponds to (German) /\textipa{rad}/ ‘wheel’ or /\textipa{rat}/ ‘rat’, speakers will guess correctly at a higher than chance level (60 to 70 per cent, according to Port & Crawford, 1999).\footnote{Among the relevant literature, we mention Dinnsen & Charles-Luce (1984), Charles-Luce & Dinnsen (1987), Wheeler (2005) for Catalan; Ernestus (2000), Warner et al. (2004), Ernestus & Baayen (to appear?) for Dutch; Dinnsen & Garcia-Zamor (1971), Mitleb (1981), Dinnsen & Charles-Luce (1984), Port & O’Dell (1985), Firoth et al. (1991) for German; Słowiński & Dinnsen (1989), Słowiński & Szymańska (1989) for Polish; Pye (1986) for Russian; Wilson (2003) for Turkish.} Port & Leary (2005) observe: “If [these words] were the same, then in a listening task you would expect 50 percent correct (pure guessing - like English too and two would show). If contrastive, one would expect at least 99 percent correct identification under good listening conditions with motivated subjects (just like [German] \textit{Bunde} and \textit{bunte} would show).” Neither is the case.

Previous phonetic results about incomplete neutralisation have suffered from methodological criticisms. For instance, it has been claimed (Fourakis & Iverson, 1984; Baumann, 1995; Manaster Ramer, 1996) that these results were obtained in a laboratory situation in which the orthography — reflecting the ‘underlying’ distinction — played a role which was larger than in ‘real life’ situations, or in labs where were based in the United States with (almost) bilingual speakers so that we could expect influence from non-devoicing English.

However, it has also been pointed out (see Charles-Luce, 1993; Jansen, 2004, for an overview) that not all experiments suffered from these problems. Furthermore, some of the more recent literature has repaired these criticisms, and still reached the same conclusions. There is thus no reason to discard them, and furthermore we also cannot say that these facts are ‘mere phonetics’, since they reflect a voicing distinction that is undoubtedly underlying, hence phonological, and a process that is also phonological at least in some cases. For instance, it interacts with phonological processes of voicing assimilation and/or it counterbleeds resyllabification across morpheme boundaries in Catalan, Polish and Dutch. And finally, it seems to me that a model of linguistic competence might still want to account for the fact that listeners are able to perceive the very subtle and subconscious phonetic details in the first place.

Furthermore, these results are not necessarily incompatible with accounts based on categorical delaryngealisation. In the literature, we can already find roughly two ways of dealing with them, both of which can be worked out in a variety of ways (cf. Bermúdez-Otero, 2006):

1. There is a phonetic paradigmatic effect (Ernestus & Baayen, to appear);
the phonetic implementation of words ending in a voiced obstruent is influenced by the fact that the same word is pronounced with a voiced obstruent in other positions in the paradigm. ‘Word-based phonetics’ (Pierrehumbert 2002) is a possible implementation of this.

2. The laryngeal contrast between voiced and voiceless obstruents is ‘enhanced’ by other features (or for instance by vowel length) before it disappears. A well-known implementation is the one by Rice & Avery (1989).

Although both these accounts are feasible, they both assume a complication of the model of phonology-phonetics interaction, if the null hypothesis is that phonology is free of phonetic substance, and inversely, phonetics blindly interprets the structures that are generated by the phonology. Under the first solution, the phonetics is not blind, but can independently access word structure. Under the second proposal, the phonology already takes care of the ‘enhancement’ that would otherwise be considered purely phonetic.

It is clear why solutions such as these have been proposed, and it is also clear why some authors have concluded from these facts that formal theories of phonology are incompatible with incomplete neutralisation. We will need a phonological (surface) structure in which devoiced segments do not carry the feature [voice] but still differ from underlyingly voiceless segments. This seems prima facie strange, but it is not a logical impossibility, given an appropriate theory of phonological representations, as we will now show.

6.2 Turbidity theory as a theory of faithfulness

The set of Optimality Theoretic constraints is usually divided into two subsets: markedness and faithfulness. Markedness constraints evaluate the output representations, and check for instance whether feet are binary, syllables have onsets, and vowels do not contain undesirable feature combinations such as [-back, +round]. Faithfulness constraints on the other hand check whether the output representation has not changed too much from the input.

In the literature, we can discern two traditions in the formalisation of faithfulness constraints. One tradition works with separate input and output representations, and correspondence relations between them; it is fair to say that most work in OT, following McCarthy & Prince (1995a), falls within this tradition. An alternative tradition, however, is based on the notion of Containment (Prince & Smolensky, 1993):

(186) Containment. Every element of the phonological input representation is contained in the output. (There is no deletion.)

In a Containment model of faithfulness, all constraints only evaluate one representation, the output. Since the input is contained in this output, it does not
6.2. Turbidity theory as a theory of faithfulness

need a separate status in the model. An important question is of course how we represent non-pronunciation in such a model. The answer which Prince & Smolensky (1993) provide is that underlying elements which are not parsed are left unpronounced. Thus, suppose that in a hypothetical language we have an input /takp/, and that this input is pronounced in isolation as [tap]. We may picture the output representation as follows:

(187)

Only /t, a, p/ are parsed into the syllable structure, and therefore into higher-order phonological structure. Constraints against ‘deletion’ in this theory of faithfulness are constraints against floating material: Parse constraints (e.g. ‘a segment needs to be parsed into a syllable’). Furthermore, it is assumed that the phonetics will only interpret the material which is parsed into the prosodic hierarchy: it is subject to a principle of Stray Erasure (Itô, 1986):

(188) Stray Erasure: The phonetics only interprets parsed phonological material.

Prince & Smolensky (1993) have implemented the idea of Containment in one specific way, the so-called Parse & Fill Model, named after the two constraints which take care of the most important aspects of faithfulness theory, deletion and insertion respectively:

(189) a. Parse: Deleted elements are ‘not parsed’ in the phonological structure
b. Fill: Inserted segments are ‘empty’

The Parse & Fill Model suffers from a number of problems (van Oosten-dorp, to appear[a]), which mainly are related to its theory of epenthesis. Inserted segments have to be ‘empty’ in order to be recognizable as inserted at all: if we would allow non-empty epenthetic segments, the constraint Fill would not be violated, and hence we would have no theoretical device to prevent gratuitous epenthesis in all languages of the world. We thus have to prevent features from being inserted or from spreading into epenthetic vowels if we use Fill. This means for instance that we can have no phonological analysis of vowel harmony to epenthetic vowels, clearly an undesirable result.

Instead of this particular implementation of Containment, (van Oosten-dorp, to appear[a]) therefore proposes an alternative implementation, which based on the principle of Consistency of Exponence, another classic principle constraining Gen in Optimality Theory:
6.2. Turbidity theory as a theory of faithfulness

This principle, assumed to restrict Gen, was explained by McCarthy & Prince (1993, 1994) in the following way:

“[Consistency of Exponence] means that the lexical specifications of a morpheme (segments, prosody, or whatever) can never be affected by Gen. In particular, epenthetic elements posited by Gen will have no morphological affiliation, even when they lie within or between strings with morphemic identity. Similarly, underparsing of segments — failure to endow them with syllable structure — will not change the make-up of a morpheme, though it will surely change how that morpheme is realized phonetically. Thus, any given morpheme’s phonological exponents must be identical in underlying and surface form.”

An important consequence of this principle is that the morphological identity of segments will be visible at the surface structure; in this way our phonological constraints can refer to them even within a monostratal model.

van Oostendorp (to appear) proposes a notation which allows us to see the effects of Consistency of Exponence, and which is based on the metaphor of colouring. It is assumed that every morpheme has its own ‘colour’ which has been provided by the lexicon and which is distributed over all segments and other material — features, mora’s, etc. — which is lexically present in that morpheme. Assume for instance that we have an input morpheme /takp/, and an output candidate which would be pronounced ad [tapi]. This candidate would look as follows in the phonological surface (for the sake of reproductional convenience, the colours are reproduced here as subscripts):

\[ \sigma \sigma \]

In this simple example, there is only one morpheme with the ‘colour’ \( \alpha \). The epenthetic segment does not have any morphological colour, which denoted here by giving \( \emptyset \) as its subscript. In terms of colours, Consistency of Exponence states that Gen cannot give colour to epenthetic material, and it cannot alter the colours of underlying material.

\[ \text{van Oostendorp (to appear)} \]

But given this notational assumption, it becomes easy to determine the status of epenthetic material by checking only the phonological output: epenthetic
material is exactly the material which does not have a morphological colour. 
Epenthetic segments thus do not have to be marked as featurally empty, since 
they are already ‘empty’ from a morphological perspective by definition. It 
now becomes possible to do away with Fill and to define constraints against 
epenthesis and deletion in a parallel fashion. Deletion means — like in the 
Parse&Fill model — that a segment is not incorporated into the phonologi-
cal structure; epenthesis means that a segment is not incorporated into the 
morphological structure.

\[(193)\]

\[\text{a. Parse-} \phi(\alpha): \text{The morphological element } \alpha \text{ must be incorporated}\]
\[\text{into the phonological structure. (No deletion.)}\]

\[\text{b. Parse-} \mu(\alpha): \text{The phonological element } \alpha \text{ must be incorporated}\]
\[\text{into the morphological structure. (No insertion.)}\]

van Oostendorp (to appear\textsuperscript{a}) argues that a large portion of the OT litera-
ture — also if it overtly is based on the theory of correspondence — has 
to assume Consistency of Exponence, for instance in order to implement 
morphologically-based positional faithfulness. If that is true, the constraints 
in (210) come for free, as it were.

However, we still have not implemented a full theory of faithfulness for 
features, given these assumptions. In particular we still need to find a way to 
describe the deletion of underlying features. Suppose we have the following 
input: (‘x’ is a root node, an X-slot, a mora, or whatever is the host for a 
phonological feature; ‘F’ is the feature itself):

\[(194)\]

Furthermore, suppose that x is not pronounced with the feature F in the pho-
netics. What should be the corresponding output representation? Obviously, 
F cannot be deleted, and the fact that it is not pronounced could be inter-
preted as it not being associated to the segment. We could thus assume that 
the following is the relevant output representation:

\[(195)\]

Yet there is a problem with the theory of feature faithfulness we have now 
developed. Suppose that Parse-F is ranked very highly in some language L. 
(194) will then be favoured over (195) in L. But how about (196)?
There is nothing in the Containment model so far which distinguishes these two representations. We need to implement the insight that also the association line needs to be preserved from the input to the output: we cannot just ‘move’ the association line from one segment to the next one. But it does not seem to make ontological sense to say that association lines can also be ‘parsed’ or ‘not parsed’ by Gen: an association line is not a phonological object on a par with features and segments, but it rather describes a relation between two phonological objects.

In order to solve this problem, Revithiadou (to appear) revises an idea originally put forward by Goldrick (2000), called Turbidity Theory. In this theory, autosegmental association lines are replaced by two types of relations:

- **projection**: an abstract, structural relationship holding between a segment and the feature (roughly equivalent to notions of ‘Licensing’).
- **pronunciation**: an output relationship that holds between the feature and the segment and describes the output realization of structure.

The representations are ‘turbid’ because the projection relations are not pronounced; they serve purely structural needs. For Goldrick (2000), there main function is the description of phenomena which are somehow opaque. For instance, in Luganda — like in many other languages — deletion of a vowel in hiatus may lead to lengthening of the vowel that is left behind.

\[
\begin{array}{l}
\text{(197)} \\
a. /\text{ka+tiko}/ \rightarrow [\text{katiko}] \text{‘mushroom’} \\
b. /\text{ka+oto}/ \rightarrow [\text{ko:to}] \text{‘fireplace (DIM)’} \\
c. /\text{ka+ezi}/ \rightarrow [\text{ke:zi}] \text{‘moon (DIM)’}
\end{array}
\]

The standard analysis of this would be that the deleted vowel leaves behind its mora which is filled by the other vowel, which thus becomes bimoraic, i.e. long. The question is, where does this mora come from? It is not necessarily underlying, since nothing is in OT: given Richness of the Base, vowels can be underlingly moraless, but still they will always have this effect. The answer of Turbidity Theory to this puzzle is that the mora is *projected* by the deleted vowel, in spite of the fact that this deleted vowel is itself not pronounced; however the mora is *pronounced* on the vowel which is itself pronounced. The turbid representation of the relevant vowel pair in (197b) is as follows:

\[
\begin{array}{c}
\mu \\
\downarrow \\
\mu \\
\uparrow \\
\text{a} \\
\end{array}
\]
The upward arrow represents the projection relation: the first mora is present because the /a/ wants it to. The downward arrow represents the pronunciation relation: the first mora is pronounced on the [o] vowel. The second mora is both projected and pronounced by this second vowel; this is the unmarked state of affairs, and there will hence be constraints requiring pronunciation and projection to match (RECIPIROCITY, see below).

Notice that Turbidity Theory presupposes Containment: this view of relations between vowels and features or mora’s does not make a lot of sense if the original vowel /a/ is lost, because in that case it also will not enforce projection of a mora. Turbidity Theory also requires a slight modification of the principle of Stray Erasure in (188), at least for subsegmental material:

\[(199) \text{Stray Erasure (Turbid version): The phonetics only interprets features that stand in a pronunciation relation to a segment in the phonology.}\]

However, notice that in its present version, Turbidity Theory does not really solve the original problem, which is how faithfulness constraints can differentiate between the representations in (194) and (196) (imagine that the lines given here are pairs of projection and pronunciation relations). In Goldrick (2000)’s original version, there is still no way to keep track of what is underlying, and what is not underlying. This is where Revithiadou (to appear)’s extension of the theory becomes important. She proposes that

we take projection lines to represent the lexical state of affairs, that is, to be part of the lexical representation of a morpheme […]

In conformity with [Consistency of Exponence], therefore, they cannot be altered by Gen.

We thus assume that projection lines represent the interface between phonology and the lexicon, whereas pronunciation lines present the interface between phonology and the phonetics: Gen can neither insert nor delete lexical material, hence it simply cannot change projection lines, but it can freely manipulate pronunciation lines (Blaho & Bye 2005 present a similar idea in terms of Correspondence Theory). Since there are constraints matching projection and pronunciation lines, however, — of the type in (200), adapted from Goldrick (2000)’s and Revithiadou (to appear)’s work — usually a feature will be sent to the phonetics on the segment on which it is specified in the lexicon. But ‘deletion’ of a feature will usually be represented by a projection line which is not interpreted as a pronunciation line:

\[(200) \text{RECIPIROCITY}^V_F (\not \in^V_F): \text{If a vowel V entertains a projection relation with a feature F, then F must entertain a pronunciation relation with the vowel V.}\]

\[
\begin{array}{c}
\text{x} \\
\uparrow \\
\text{F}
\end{array}
\]

\[(201)\]
6.3 The analysis of final devoicing

In the analysis of final devoicing, we have to make a number of analytical choices, for instance regarding the representation of the laryngeal contrast — which features are involved, are these features binary or monovalent, etc. —, or the precise context — end of the syllable, end of the word, etc. (see the contributions to van der Torre & van de Weijer [in press] for an overview of such possibilities for Dutch). It is also possible that different analytical tools are needed for different languages showing final devoicing.

The precise formulation of the relevant constraints obviously falls outside of the scope of the study of faithfulness and hence of this paper: these will be pure markedness constraints. Our proposals are independent of them. In order to make our proposal maximally explicit, we will use the following constraint (Lombardi [1991, 1999] for more details):

\[(202) \text{FINDEV} \]

\[\text{[voice]} \text{ cannot entertain a pronunciation relation with an obstruent in the coda.}\]

Furthermore, I will assume for the sake of concreteness that [voice] is a monovalent feature, contra Wetzels & Mascaro (2001): an obstruent has a [voice] feature, or it does not have it. In other words, for the Dutch adjective kwaad ‘angry’, ending in an underlying [voice] there are two relevant potential realisations: one in which the projected feature is also pronounced, and one in which the pronunciation relation is not realised. The latter is the devoiced representation:

\[(203) \]

\[\begin{array}{c|c}
\text{a. [kwaːd]} & \text{b. [kwaːd]} \\
\hline
\text{k w aː d} & \text{k w aː d} \\
\text{\uparrow} \text{ voice} & \text{\uparrow} \text{ voice}
\end{array}\]

FINDEV will prefer the latter one, but the former will be preferred by \(\text{RECIPROCITY}^V_F\), so that we have a miniature typology:

\[(204) \]

\[\begin{array}{c}
\text{a. FINDEV} \gg \text{RECIPROCITY}^V_F: \text{Final Devoicing (Catalan, Dutch, etc.)}
\end{array}\]
The reader will have noticed already that given the phonological apparatus introduced for independent reasons so far we actually have a phonological surface structure which is transcribed more accurately as \([\text{kwa:}d]\) than as \([\text{kwa:t}\])\). An underlyingly voiceless segment will have a different phonological representation, viz. one in which [voice] is completely absent.

Given the interpretation given to Stray Erasure in (199) above, this phonological difference does not have an effect in the phonetics. What the facts of incomplete neutralisation in section 6.1 suggest, then, is that this interpretation may be too rigid. In some cases, also features which are only projected play a role in the phonetic interpretation. For instance, if the length of the preceding vowel is a cue to voicing, this lengthening may be triggered slightly by projected voicing features as well. Of course, it is required that the phonetics is no longer categorical, but gradient: it will interpret the phonological \([\text{d}]\) in a number of very subtle ways, whereas phonological \([\text{d}]\) and \([\text{t}]\) get a more stable — yet also still gradient — interpretation. Yet I believe there is nothing exotic about the assumption that the phonetics is no longer a categorical system.

Notice that this model is able to maintain a classical view of the relation between phonology and phonetics. Phonology is free of phonetics: the difference between projection and pronunciation lines was introduced for completely independent reasons, viz. to describe phonological opacity, and it is the only phonological device we need to describe the difference between devoiced and voiceless. There is no introduction of ‘enhancing’ features in the phonology. Phonetics, inversely, still does not need to ‘know’ anything apart from the phonological surface structure: it does not need to refer to the lexicon directly, since it gets all the lexical information it needs from the phonological surface.

### 6.4 Concluding remarks

The debate about ‘functionalism’ vs. ‘formalism’ has been revived in recent years, but it has been conducted in a manner which might be called one-
sided: in a number of works, ‘functionalists’ have been pointing out facts which they consider to be problematic for formalist views of phonology.

We have to point out that the reverse — finding facts which are hard to account for in a functionalist view is inherently more difficult, since such a view seems inherently less restrictive, at least in as far as we do not have an explicit theory of possible ‘functions’ of language, and of possible ways in which these functions can be worked out. The article ‘Against formal phonology’ by Port & Leary (2005), cited earlier, for instance, does not present a detailed alternative proposal: it attacks a school of thought without replacing it with something which has the same level of explicitness. In the absence of such an alternative, it is hard to even think of potential data which would falsify the ‘functionalist’ view (what is a logically possible fact which can be ‘proved’ dysfunctional?).

Given this, a typical ‘formalist’ response to the functionalist challenge has been to point out problems with the data, or to ignore those data completely. I have argued in this article that the evidence for incomplete neutralisation seems too convincing to completely ignore it; it sheds light at least on the interface between phonology and phonetics. The fact that most speakers are not able to consciously hear the difference between a devoiced stop and an underlyingly voiceless one, does not count as sufficient evidence: phonology is not (just) about conscious judgements, but about any kind of way in which categorical distinctions play a role in linguistic behaviour. Just delegating these facts to the phonetics will mean that the phonetics gives access to many types of information which is purely lexical — and in a way this may vindicate the functionalist position: if the phonetics can see all this information anyway, why would we need the phonology to manipulate them as well?

Formal phonologists thus need to take these facts seriously, and try to incorporate them into their model of phonology. The more conservative approach is to not give up the whole enterprise of formal analysis in the face of a few problematic data, but try to incorporate the apparent recalcitrant facts without making the system too flexible so as to be no longer falsifiable. An additional problem is that functionalists often argue against fairly simplistic formal models, such as those presented — almost 40 years ago — by Chomsky & Halle (1968). In this article I have shown how the facts of incomplete devoicing can be incorporated into a purely formal as well as fairly elegant and simple theory of phonological representations in which we can represent that two categories ([t]) and ([t]) are “neither discretely different nor are they the same”, in the words of Port & Leary (2005). If we accept that it is better to have an explicit theory of a phenomenon, and that simpler theories are preferable over more complex ones, this means that incomplete devoicing seems an argument in favour of formal phonology rather than against it.


7 Dutch Diminutives and the Question Mark

7.1 Monostratalism and the null parse

Within a Correspondence framework of faithfulness, there are roughly four ways of treating the problem of ineffability:

1. The ‘paradigmatic solution’: the Generator function does not generate an individual form, but a paradigm. Ineffability of an individual form means that this particular form is not generated within the paradigm. This type of solution is defended in this volume (in various forms) by Bat-El, Rice, and Törkenczy & Rébrus.

2. The ‘null parse’ solution: the Generator function generates a candidate in the phonology which does not have a phonetic interpretation, and this is selected as the winner in certain cases; this type of solution is defended in this volume by McCarthy and Wolf.

3. The ‘control’ solution: the Generator and Evaluator function conspire to create a (pronounceable) candidate, but a grammatical component outside of the OT system then blocks this candidate. This solution is defended by Orgun & Sprouse in this volume.

4. The ‘divergent meaning’ solution: we generate a phonologically well-formed form, but one which does not have the intended semantics; the form is therefore unusable. This is a solution that is related to the control component solution, except that ‘Control’ does not consist of grammatical constraints, but is a (pragmatic) self-monitoring device, aimed to check whether one actually pronounces what is intended. This solution is basically the one proposed by Legendre and by Vogel in this volume for syntax, and will be defended here for phonology.

This article builds on a theory of faithfulness that is different from Correspondence Theory, viz. one which I call Consistency of Exponence (CoE) Theory, i.e. the version of faithfulness originally proposed by Prince & Smolensky (1993). The idea of the Null Parse was first brought up by these authors within the context of this theory of faithfulness, albeit not in a very formal way. Containment was soon replaced as a theory of input-output relations by Correspondence Theory (McCarthy & Prince 1995a), but if I am not mistaken, interest is currently reviving. One important advantage of CoE Theory over Correspondence is that it is theoretically more restrictive, and therefore makes more fine-grained empirical predictions.

Formally, the main difference between CoE Theory and Correspondence Theory is that the former offers a purely monostratal view of linguistic representations. All Containment based constraints, those checking faithfulness as well as those checking wellformedness, only consider one level of representation — the output. We may posit an input to the Gen function, but the constraints of Con are not aware of it, and hence the input representation
does not exist for the Eval function. Within Correspondence Theory, on the other hand, there are typically two levels of representation — the input and the output — and constraints refer to properties of one of these levels, if they are wellformedness constraints, or to the correspondence relations between them, if they are faithfulness constraints: both input and output exist both for Gen and for Eval.

If we have only one level of representation, such as in CoE Theory, we still need to be able to see deleted material, otherwise there is no way to penalize it: we need constraints against deletion, otherwise all deletion will come for free, and since we only consider one representation ‘deleted’ material still needs to be present in that representation. In classical CoE Theory of Prince & Smolensky (1993) this was accomplished by assuming (i) that all input material would be contained in the output (this was called the principle of Containment, hardwired into the Generator function), but that (ii) the material which is not pronounced is not parsed in the phonological structure: unpronounced features are left unassociated to segments, unpronounced segments are not incorporated into the syllable structure. The phonetics, coming after the phonology, only considers the phonologically parsed material and applies ‘stray erasure’ to the rest. The phonological derivation thus is monotonic whereas the phonetics is not necessarily.

The null parse solution is more problematic for CoE Theory than for Correspondence Theory: if all the material is present in the output anyway, why would we ever choose to not to pronounce any part of that material? The second possible theoretical option mentioned above, the one couched in terms of paradigms, is basically impossible to formulate in terms of CoE theory. Paradigmatic relations may be described in terms of Correspondence, but it does not make sense to try to describe the in terms of CoE, since that would involve putting all members of a paradigm into one phonological representation. This means that the number of theoretical options is reduced by half if we accept CoE, a restrictiveness which might itself considered to be a positive result: we either should choose the control solution or the divergent meaning solution.

The bigger popularity of Correspondence Theory may account for the fact that almost all of the studies on the issue of ineffability — for instance, all other phonological chapters in this book — are couched in terms of Correspondence Theory. As a matter of fact, the only Containment attempt at a solution of which I am aware is Prince & Smolensky (1993, p. 51–52)’s suggestion that the null parse is the form which does not get morphological structure assigned to it. The input of the phonological derivation for some form can be assumed to be \{ root, affix \}, i.e., an unordered set. Gen can decide to assign morphological structure to this set, e.g. [[root] affix] (violating a constraint MPARSE), or it can decide to leave the whole structure as it is morphologically unparsed. \{ root, affix \}, however, is something that will not fit into a syntactic structure, and this will cause the crash. See the introduction
to this volume for a more detailed explanation.

The discussion of this issue is rather informal in Prince & Smolensky (1993), and it therefore leaves several questions open. For instance, it is not exactly clear where the crash takes place that leads the underlying material to be unpronounced. One way to interpret Prince & Smolensky (1993)’s suggestion is that the null parse is assigned in a morphological module, which is then input to the syntax; here, the word becomes ‘uninterpretable’ however. Yet, if syntax is also an OT system, which should be the default option, it should not crash on receiving an uninterpretable input, but rather do something else (such as making it interpretable). This is of course even more so if phonology, morphology and syntax work in parallel. It thus still is not completely clear how ineffability works under Prince & Smolensky (1993)’s assumptions. More discussion of this issue will be provided in section 7.2.

This article aims to fill this gap in the theoretical literature for two reasons. In the first place, I consider CoE Theory a serious alternative to Correspondence, as I will explain in section 7.2, but as such it needs a worked-out approach of ineffability, among other things, which can compete with the best Correspondence Theory alternatives. In the second place, I believe that CoE Theory puts into a special light an important aspect of all cases of ineffability seen in the literature: that the relation to morphology and to the input plays an important role in it. The results of this investigation may therefore also be of interest to students of ineffability within Correspondence Theory.

As a matter of fact, I will present my solution to the ineffability problem within Containment in two steps. In section 7.4, I will develop a ‘radical’ solution which should be able to compete with Correspondence-based alternatives. However, I will show that often the ‘ineffable’ cases are actually not completely impossible, and may grow more acceptable to speakers when they hear them repeatedly. In section 7.5, I therefore set a few additional steps, first embedding my approach into an accepted theory of language variation to be able to describe this pronounceability of ‘ineffable’ material, and secondly, explaining why lexicalisation helps making a form more acceptable.

Next to these theoretical goals, this article also serves to add some new facts to the discussion about ineffability. These facts are from (Standard) Dutch: it is shown in section 7.3 that diminutives are difficult to form on the basis of certain proper names and nouns, depending on their phonological shape. The particular relevance of these facts is that diminutive formation seems ‘difficult’ rather than impossible: while most native speakers agree that these forms are funny, this does not mean that everybody rejects them outright. As a matter of fact, many informants seem to agree that one could use these forms if forced to, even if they sound a little funny, and a corpus search reveals that they are indeed used, albeit less frequently than forms with a similar meaning but a different phonological shape. In other words, these forms have a question mark rather than an asterisk.
The question mark poses of course a problem to OT which is of the same nature as ineffability. An OT grammar only knows absolute judgements: there is one candidate output which will win the competition — and therefore be grammatical —, and there are many candidate outputs which will be defeated (and therefore get a star), but there is nothing in between. We argue in section that the question mark indeed is an extragrammatical category, but one which can nevertheless be understood in terms of our containment based solution and a theory of language variation.

An important difference between the question mark and ineffability is that the latter is a problem mainly for OT and not for other models, such as classical, rule-based SPE-type phonology or Government Phonology: most of these alternatives assign absolute markings to all representations, which can result in some combination of morphemes to be simply impossible. However, the question mark poses serious problems for virtually every (formal) theory of grammar, and maybe even especially so for the theories just mentioned, with their absolute markings. Given the fact that the study of variation — comparing different forms for one input — is relatively well-developed within OT, the existence of question marks will be argued to actually be an argument in favour of such a model of grammar.

7.2 Containment

Containment and Correspondence

In the OT literature, we find roughly two ways of evaluating the relationship between input and output. One is monostratal, and exemplified by CoE Theory: the constraints can see only one representation, the output. The other one is multistratal — usually bistratal: the constraints can see and evaluate input and output and (correspondence) relations between them.

As already pointed out above, within a Containment model we need to be able to see in the output representation which elements are inserted and which elements are deleted. If we would not be able to differentiate an inserted vowel from an underlying vowel at the surface level, there would be no way in which we could block massive epenthesis; and if deleted elements do not leave some trace in the surface structure, there would be no way of preventing massive deletion. The latter point means that input mate-

---

41 The fact that CoE Theory is technically monostratal does not imply that it would be incompatible, for instance with Stratal OT (Kiparsky 2000). The ‘monostratal’ aspect of it involves the relation between input and output, but of course, we can still make the output of one level the input of the next level.

42 In most versions of Correspondence Theory, there are no constraints evaluating the input structure, but the reason for this is that such constraints would not extend a lot of empirical power, since all candidates in a tableau have the same input, and thus will have the same violations for such constraints. There is no logical ban on constraints on inputs.
rial should still be present in the output representation; hence the generator function is subject to a principle of Containment:

\[(205)\] Containment. Every element of the phonological input representation is contained in the output. (There is no deletion.)

Faithfulness constraints are formalized by Prince & Smolensky (1993) in the following way. They assume that ‘deletion’ means that elements are ‘not parsed’ into the phonological structure, as outlined above. ‘Inserted’ segments are supposed to remain empty — there is no insertion of features —, and this is how the phonology can recognize them.

\[(206)\] a. PARSE: All elements should be ‘parsed’ in the phonological structure (no deletion.)

b. FILL: Do not allow empty elements. (No insertion.)

An advantage of a Containment approach to faithfulness is that it is theoretically parsimonious: it does not refer to any device which is not needed independently. For instance, the highly abstract correspondence relations are not necessary. Furthermore, the PARSE and FILL families of constraints supposedly are necessary outside the theory of faithfulness proper: we need to say that syllables should be parsed into feet, regardless of whether unparsed syllables are pronounced or not; we thus need PARSE constraints or some equivalent anyway. Similarly, we will want to prevent phonologically ‘empty’ segments, for instance because they do not seem to occur in all languages; we thus need FILL constraints or some equivalent anyway, even if we subscribe to Correspondence Theory.

This thus poses a problem for many of the proposed alternatives within other phonological frameworks of dealing with faithfulness: we may stipulate that we no longer use PARSE-C or FILL-V, but then we will still need to say something about consonants that are not attached to syllable nodes on the surface, or vowels that do not have any vocalic feature content. Occam’s razor thus seems to run against introducing correspondence relations to accomplish something that can already be done. In that sense, Containment comes very close to the null hypothesis regarding faithfulness theory, given the other theoretical assumptions that were made in Prince & Smolensky (1993), McCarthy & Prince (1993).

However, CoE Theory also has a few problems. One of these is that it seems less well-equipped to deal with interrepresentational relations beyond input-output relations, such as output-output relations (Burzio 1994, 1998, 2000, 2003; Benua 1997; Kager 1999), relations between candidates in a tableau (Sympathy Theory McCarthy 1999) and, most uncontroversially, for the relations between bases and reduplicants (McCarthy & Prince 1995a).
However, alternatives for each of these presumed extensions of Correspondence Theory exist, which do not use Correspondence relations. For instance, instead of output-output relations we can use Stratal OT (Kiparsky, 2000) or Derivational OT (Rubach, 2003); for Sympathy Theory there is a spade of alternatives, and as a matter of fact, CoE Theory seems particularly suitable to deal with at least certain types of opacity (van Oostendorp, to appeara); and instead of Base-Reduplicant Correspondence, we can assume that copying in reduplication is governed by the morphology rather than the phonology (Inkelas & Zoll, 2005).

Other problems have to do with the specific implementation of the idea of Containment with PARSE and FILL rather than with that idea itself. In particular the theory of epenthesis implied by the FILL constraints is very problematic. It should be the case, for instance, not just that features cannot be inserted, but features should also not be allowed to ever spread to an epenthetic vowel. If ever a vowel which is inserted by Gen would be able acquire phonological features in whatever way, the epenthetic vowel would no longer be empty, hence it would no longer violate FILL, and the doors would be open to massive unpunalyzed epenthesis, at least in languages that allow some minimal form of vowel harmony.43

For this reason, van Oostendorp (2005c, to appeara) develops an alternative implementation of CoE Theory, which evades these problems. This alternative is based on the notion of Consistency of Exponence, another classic principle of Optimality Theory which has not been taken sufficiently seriously in my view:

\[(207) \text{Consistency of Exponence}\]

“No changes in the exponence of a phonologically-specified morpheme are permitted.” (McCarthy & Prince [1993, 1994])

This principle, assumed to restrict Gen, was explained by McCarthy & Prince (1993, 1994) in the following way:

“[Consistency of Exponence] means that the lexical specifications of a morpheme (segments, prosody, or whatever) can never be affected by Gen. In particular, epenthetic elements posited by Gen will have no morphological affiliation, even when they lie within or between strings with morphemic identity. Similarly, underparsing of segments — failure to endow them with syllable structure — will not change the make-up of a morpheme, 43As a matter of fact, the existence of vowel harmony and other types of feature spreading poses problems for the PARSE and FILL model. If spreading can occur, how do we prevent it from happening everywhere in every language? The only reasonable answer to this is: by way of faithfulness constraints against insertion of association lines. But how can we formalize that if constraints against insertion are FILL constraints — what does it mean to say that an association line is empty?
though it will surely change how that morpheme is realized phonetically. Thus, any given morpheme’s phonological exponents must be identical in underlying and surface form.”

An important consequence of this principle is that the morphological identity of segments will be visible at the surface structure; in this way our phonological constraints can refer to them even within a monostratal model. van Oostendorp (2005c) proposes a notation which allows us to see the effects of Consistency of Exponence, and which is based on the metaphor of colouring. It is assumed that every morpheme has its own ‘colour’ which has been provided by the lexicon and which is distributed over all segments and other material — features, mora’s, etc. — lexically present in that morpheme. Assume for instance that we have an input morpheme /takp/, and an output candidate which would be pronounced ad [tapi]. This candidate would look as follows in the phonological surface (for the sake of reproductional convenience, the colours are reproduced here as subscripts):

\[
\sigma \quad \sigma
\]

\[
t_\alpha \quad d \alpha \quad k_\alpha \quad p \alpha \quad i_\emptyset
\]

In this simple example, there is only one morpheme with the ‘colour’ \(\alpha\). The epenthetic segment does not have any morphological colour, which denoted here by giving \(\emptyset\) as its subscript. In terms of colours, Consistency of Exponence states that Gen cannot give colour to epenthetic material, and it cannot alter the colours of underlying material.

\[(209)\quad \text{Consistency of Exponence (Colour-based version). Gen cannot change the morphological colour of any phonological element.}\]

But given this notational assumption, it becomes easy to determine the status of epenthetic material by checking only the phonological output: epenthetic material is exactly the material which does not have a morphological colour. Epenthetic segments thus do not have to be marked as featurally empty, since they are already empty from a morphological perspective by definition. It now becomes possible to do away with FILL and to define constraints against epenthesis and deletion in a parallel fashion. Deletion means — like in the PARSE&FILL model — that a segment is not incorporated into the phonological structure; epenthesis means that a segment is not incorporated into the morphological structure.

\[(210)\quad \begin{align*}
a. \quad \text{PARSE-\(\phi\)(\(\alpha\))}: & \quad \text{The morphological element \(\alpha\) must be incorporated into the phonological structure. (No deletion.)} \\
b. \quad \text{PARSE-\(\mu\)(\(\alpha\))}: & \quad \text{The phonological element \(\alpha\) must be incorporated into the morphological structure. (No insertion.)}
\end{align*}\]
Consistency of Exponence has recently come under attack (Walker & Feng, 2004; Łubowicz, 2005a), but van Oostendorp (to appeara) argues that these attacks are not very convincing, and as a matter of fact that a large literature has to assume Consistency of Exponence, for instance in order to implement morphologically-based positional faithfulness. If that is true, the constraints in (210) come for free, as it were.

The principle of Containment can be seen as a lemma of Consistency of Exponence under the original definition of that constraint in (207): deletion of underlying segments of a morpheme \( m \) would mean changing the exponence of \( m \). However, the equation no longer holds under the colour-based definition of Consistency of Exponence given in (209). Consistency of Exponence just states that if segments are preserved, they will keep their original morphological colour. This no longer implies that they have to be preserved, just like the principle of Containment does not say anything anymore about the colouring of phonological material. The two principles thus have become logically independent, something which we will use later on.

### Ineffability in classical CoE Theory

It thus seems useful to take the theoretically parsimonious alternative that CoE poses to Correspondence Theory seriously. This means, however, that we have to provide alternative analyses for theories which have been proposed within Correspondence Theory, such as the analyses of ineffability that are developed in the other chapters of this book.

In order to do this, it is useful to briefly consider Prince & Smolensky (1993)'s discussion of these issues. As a matter of fact, these authors entertain two slightly distinct possibilities. In the first place, we might assume that in some cases it is better to not assign e.g. a foot or a phonological word label at all. If Latin does not have monomoraic words, this effect may be due to the interaction between \( \text{FTBIN} \) and the constraint \( \text{LX} \approx \text{PR} \), requiring every lexical word to correspond to a phonological word (this fulfilled a function in the theory which roughly corresponds to that of \( \text{ALIGN}(\omega, X_0) \)).

If \( \text{FTBIN} \gg \text{LX} \approx \text{PR} \), it may be better not to assign any phonological structure at all in order to prevent monomoraic forms from surfacing:\footnote{A very straightforward, and probably more correct, account for this fact is later on also provided by Prince & Smolensky (1993): an underlying form \text{/r̺e/} would be subjected to vocalic lengthening to \text{[r̺e]}. But if this vowel always surfaces as lengthening, Lexicon Optimisation will cause the long vowel to also be underlying.}
Candidate (211b) represents an attempt to parse the structure into phonological constituency, but this results in a non-binary foot. It is then better to not parse the whole structure at all. The result of this is that FTBIN is satisfied vacuously, since there is no foot to violate it. Candidate (211a) will thus win, and handed over to the phonetics. But since the phonetics will only pronounce the material affiliated to a phonological word, the pronunciation of this winning item will be zero.

This solution works well to describe certain Morpheme Structure Constraints (MSCs) for short words (‘stems cannot consist of only a short vowel’), viz. precisely in those circumstances in which the relevant markedness constraint, FTBIN in this case, dominates all faithfulness constraints. However, cases of this type are rare and furthermore, MSCs may be derivable by other means (see for instance McCarthy, 1998, for a paradigmatic approach). Furthermore, this solution does not work in most more complicated cases, where it is not so clear why we could not parse at least a part of the relevant structure. Consider the comparative form of the English adjective violet. It is well known that the comparative suffix -er can only be attached to Minimal Words of the size of one foot. Now suppose that we input {violet, er} to the grammar: what is the output of that particular set of underlying morphemes? From the point of view of phonology, there can be nothing against an output \([\omega(vio)F \text{let}F]\), since this is a well-formed structure elsewhere in the language, viz. in the simple form of the adjective; so it is not clear why the phonological Null Parse would ever win.

In order to be able to deal with such facts, then, Prince & Smolensky (1993, p. 53) introduce another type of parsing constraint, MPARSE. The idea is that in the input words do not have any morphological structure:

On this view, then, the underlying form of an item will consist of a very incompletely structured set of specifications which constrain but do not themselves fully determine even the morphological character of the output form. These specifications must be put in relation, parsed into structure, in order to be interpretable.

The constraint MPARSE (M for morphology) requires that in the output all the relevant structure has been assigned. Furthermore,

Failure to achieve morphological parsing is fatal. An unparsed item has no morphological category, and cannot be interpreted, either semantically or in terms of higher morphological structure.

I have already pointed out in section 7.1 above that it is not clear how this informal suggestion can be implemented precisely. If semantics or higher-
order morphosyntactic structure also work as an OT system, they still cannot crash. If we assume a serial model — if we apply lower-level morphology and phonology first, and feed the results of this to ‘higher morphological structure’ or semantics, these modules will still try to get a result out of this. The same is true if all modules work in parallel. The only realistic answer therefore should be that it is eventually interpretative modules outside of (OT) grammar such as phonetics or pragmatics which will cause the crash.

More importantly, however notice that the OT system has to be set up in a way which is not fully compatible with the assumption of Richness of the Base, in order for the analysis to work. It is crucial that the input is morphologically and phonologically underspecified. If we would allow inputs which are already parsed underlyingly, these inputs could presumably not be bested by a Null Parse, especially under an architecture of the grammar in which Containment plays a role, i.e. in which no underlying material can be thrown away. It seems therefore necessary to provide a more precise theory of the Null Parse within CoE Theory. This will be attempted in the sections 7.4 and 7.5 below, using the concepts of Coloured Containment. But before doing that, I will first provide some data which shed new light on this issue.

7.3 Dutch diminutives

The crucial data

The formation of diminutives has been the object of intensive study within Dutch linguistics for a long period of time (see Trommelen, 1982; Kooij, 1982; van der Hulst, 1984; Booij, 1995; Gussenhoven & Jacobs, 1998; van Oosten-dorp, 2000; Botma & van der Torre, 2000; van de Weijer, 2002; for some more recent contributions). The process is very productive and can affect all nouns and personal names in principle:

<table>
<thead>
<tr>
<th>base form</th>
<th>diminutive form</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>man</td>
<td>man-otjɔ</td>
<td>man</td>
</tr>
<tr>
<td>maan</td>
<td>maan-tjɔ</td>
<td>moon</td>
</tr>
<tr>
<td>raam</td>
<td>raam-pjɔ</td>
<td>window</td>
</tr>
<tr>
<td>dak</td>
<td>dak-jo</td>
<td>roof</td>
</tr>
<tr>
<td>Orhan</td>
<td>Orhan-otjɔ</td>
<td>(name)</td>
</tr>
<tr>
<td>Geraldine</td>
<td>Geraldin-tjɔ</td>
<td>(name)</td>
</tr>
<tr>
<td>Ralf</td>
<td>Ralf-jo</td>
<td>(name)</td>
</tr>
</tbody>
</table>

The diminutive displays a substantial amount of allomorphy, and this is the focus of most phonological work, since this allomorphy is largely determined by prosodic factors. For instance, if a noun ends in a lax vowel plus a sonorant (man), the diminutive takes the shape -otjɔ, whereas if the noun ends in a tense vowel plus a sonorant (maan), the suffix takes the shape -otjɔ.
Yet a footnote in a recent study in terms of Optimality Theory is particularly interesting. van de Weijer (2002) notes — as far as I can tell for the first time in the literature — that certain nouns cannot take a diminutive suffix. After having consulted the intuitions of a number of native speakers of Standard Dutch, I conclude that this concerns in particular those ending in -/t/ or /d/.

Note that in the case of names, there is the option of also adding the alternative diminutive suffix -ke, which seems to give a more acceptable result for most speakers. This applies only to names, since nouns never get this suffix — they sound dialectal if this is attempted.

If C in -/Ca/ is a different segment, judgments vary, reflecting possibly subtle differences in individual grammars, or the fact that words of this type are almost all proper names and therefore display a slightly different behaviour:

This is the ineffability fact that will be at the focus of our attention in this paper. In order to properly understand it, we first need some background on

---

The German diminutive sometimes also leads to ineffability, albeit of a completely different kind, having to do with umlaut. Cf. Fery (1994); Fery & Farselow (2003) and section 7.6.
the morphophonological status of the diminutive suffix.

**A sketch of the analysis of diminutives**

Although this has been phrased in different ways in the literature, the prosodic sensitivity of the diminutive suffix is usually taken to mean that the diminutive needs to be added to a minimal word of a certain shape. In other words, the following constraint applies to the diminutive suffix:47

\[(215) \text{SUFF-TO-PRWD: The base of suffixation is a bisyllabic word.}\]

Some scholars (e.g. van Oostendorp, 2000) assume that tense (long) vowels in Dutch always head an open syllable. The consequence of this is that the final consonant of *maan* appears in a separate position, e.g. the onset of an empty-headed syllable. We thus have the following structures for *maan* and *man* respectively:

\[(216)\]

a. *maan* ‘moon’ [maːn]  
\[\sigma \sigma\]  
\[m a: n 0\]

b. *man* ‘man’ [mən]  
\[\sigma\]  
\[m a n\]

According to SUFF-TO-PRWD, we can affix the diminutive suffix to (217a) without any problems, but — assuming that monosyllabic units do not form minimal words — something needs to be added to the structure in (217b). This may be a schwa, providing the proper prosodic basis for the form in question:

\[(217)\]

a. *maantje* ‘moon’ [maːntjə]  
\[\omega\]  
\[\frac{\omega}{\sigma}\]  
\[\frac{\frac{\frac{\frac{\frac{\frac{m a: n 0}{n \sigma}{tj \sigma}}}{}{}}{}}{}}{}}\]

b. *man* ‘man’ [məntjə]  
\[\frac{\omega}{\sigma}\]  
\[\frac{\frac{\frac{\frac{\frac{\frac{m a n n \sigma}{n \sigma}{tj \sigma}}}{}{}}{}}{}}{}}\]

Given this analysis, we can now wonder about the morphological affiliation of the schwa: it does not make sense to say that it is part of an allomorph of the stem, because that would involve positing lexical allomorphy for a (large) natural class of forms; all words ending in a short vowel and a sonorant would have an allomorph with a schwa. There are therefore three options:

---

47This analysis loosely follows van de Weijer (2002).
1. The schwa has its own morphological status as a separate (binding) morpheme. Such an analysis is possible, because there is independent evidence that Dutch uses schwa in this function. In compounds of a certain shape, we find schwas inserted which are probably not epenthetic: e.g. *vrouw-Ə-lichaam ‘female body’ from *vrouw ‘woman’ and lichaam ‘body’. As far as I can see, it would however be a novelty in the literature to assume that this binding schwa also shows up in the diminutive, and since nothing seems to bear on the issue for the present discussion, I have decided to assume the standard analysis. See Kooij (1982) for further discussion.

2. The schwa is phonologically epenthetic, and stays without morphological affiliation. Something similar applies to this: Dutch has phonological epenthesis of schwa (van Oostendorp, 2000), but this does not normally occur between two consonants with the same place of articulation, such as the two coronals in *mannotjo; for instance it does not apply in a word like hand ‘hand’, although it does apply in a word like kalk ‘chalk’.

3. The schwa is part of an allomorph of the diminutive suffix. This seems then the most reasonable solution. A solution for the Dutch diminutive in terms of allomorphy is also suggested in Booij (2002): the suffix can also take the shapes -kja, -pjə or -ja, and although each of these forms is phonologically conditioned, it is hard to relate them to productive processes of Dutch phonology — for instance, voiceless stops do not otherwise assimilate to neighbouring segments.

It has often been observed that the schwa suffix does not show up after a short vowel followed by an obstruent: the diminutive of *dak ‘roof’ is dakje, not *dak-Ə-tje. Botma & van der Torre (2000) argue that the reason for this is that obstruents get a different prosodification from sonorants in Dutch: obstruents are always in a separate syllable, even if they are preceded by a short vowel.

\[ \begin{array}{c}
\omega \\
| \omega \\
| \Omega \\
| \sigma \\
| \sigma \\
\langle d a k k j o \rangle \end{array} \]

(218)

This solution raises the question why the constraint SUFF-TO-PRWD forces the selection of a special allomorph in the case of sonorants rather than using the option of an empty segment. A possible answer is that this is a cyclicity effect — in terms of Correspondence Theory, it would be an effect of Output-
Output correspondence. If we assume that maan and dak are bisyllabic already before attachment of the suffix, we do not have to change anything to adjust to. Yet bal needs to be changed, in order to form a minimal word, and for this reason we choose to insert a schwa.

**Why the ungrammatical forms are wrong**

Given the rough structure of the diminutive forms, we will now try to find the reasons why diminutives are considered to be ungrammatical if the stem ends in an obstruent. One obvious correlate is the observation that two schwas in a row are dispreferred in Dutch. We will informally call this constraint *\text{\textcopyright} (van Oostendorp, 2000):

\[(219)\quad \text{*\textcopyright}: \text{Two consecutive syllables should not contain schwa.}\]

This constraint is responsible for other forms of allomorphy as well. For instance, the agentive suffix -\text{\textcopyright}er takes an allomorph -\text{\textcopyright}aar if the stem ends in a syllable headed by schwa (220a). Also, the allomorphy between the productive plural suffixes -\text{\textcopyright}on and -\text{\textcopyright}s seems to be governed by it at least partly (220b):

\[(220)\]
\[\begin{align*}
\text{a.} & \quad \text{denk-\text{\textcopyright}r ‘thinker’, } \text{lop-\text{\textcopyright}r ‘walker’, } \ldots \\
\text{b.} & \quad \text{piek-\text{\textcopyright}r ‘thinker’, } \text{wandsel-\text{\textcopyright}r ‘walker’, } \ldots
\end{align*}\]

\[\text{a.} & \quad \text{berg-\text{\textcopyright}n ‘mountains’, filosof-\text{\textcopyright}n ‘philosophers’, } \ldots \\
\text{b.} & \quad \text{heuvel-s ‘hills’, denkor-s ‘thinkers’, } \ldots
\]

The constraint is not absolutely surface true for Dutch, however. Certain suffixes, such as the comparative, verbal inflection, or one allomorph of the nominal plural, do not seem to be subject to it. We can even combine certain affixes to create longer sequences of schwas as in (221d):

\[(221)\]
\[\begin{align*}
\text{a.} & \quad \text{groot-\text{\textcopyright}r ‘bigger’, } \text{edel-\text{\textcopyright}r ‘nobler’, } \ldots \\
\text{b.} & \quad \text{lop-\text{\textcopyright}n ‘walk (Pl)’, } \text{wandsel-\text{\textcopyright}n ‘walk (Pl)’, } \ldots \\
\text{c.} & \quad \text{kind-\text{\textcopyright}n ‘children’}
\end{align*}\]
\[\text{d.} & \quad \text{kind-\text{\textcopyright}-lak-\text{\textcopyright}-o ‘more childish’ (child+binding morpheme+adjectivizing morpheme+comparative+agreement)}
\]

What is more, even the diminutive suffix itself does not always obey the constraint, viz. in those cases in which a special allomorph is chosen after a stem ending in a lax vowel plus a sonorant consonant:

\[(222)\quad \text{bal-\text{\textcopyright}jo ‘little ball’}\]
Notice also that the effect only holds if the schwa is the last segment in the word, not if it is followed by a consonant: words such as *vador*ˈtɔːr*ˈtɔːr* ‘little father’, *têgal*ˈtɔːr*ˈtɔːr* ‘little tile’ or *bez*ompjˈtɔːr*ˈtɔːr* ‘little broom’ are unobjectionable.\(^{48}\)

The constraint \(^*_{\omega}\) in (219) should therefore be highly sensitive to morphological information: it clearly is part of lexical rather than of postlexical phonology. Furthermore, it seems that we are dealing with a type of derived environment effect. I will provisionally assume, therefore, that the relevant constraint is formulated as follows:

\[
(223) \quad \!_{\omega}\!_{\omega}: \text{Two consecutive syllables with a different morphological colour should not contain schwa.}
\]

The constraint should probably more precise in order to deal with the facts mentioned above, but the present format will be sufficient for our purposes. (See section 7.5, footnote\(^{55}\) for a suggestion that we can do away with the colours in this constraint.)

Furthermore, another factor plays a role, viz. the quality of the preceding consonant. The fact that the judgements are more equivocal in case the last consonant of the stem is a coronal stop must have something to do with the fact that the diminutive suffix itself also starts with a coronal stop. Now there is some evidence that a constraint of the following type also holds an effect on the Dutch lexicon (van Oostendorp, 2000):

\[
(224) \quad \!_{C}\!_{C}: \text{A schwa should not be surrounded by two identical consonants.}
\]

Again, it does not seem possible to create structures violating this constraint by morphological derivation or inflection.\(^{49}\) For instance, the comparative suffix and the agentive suffix both are -\(\mathord{\text{or}}\). They can both be attached to stem ending in an \(\mathord{\text{/r/}}\), but in this case, they take the allomorph -\(\mathord{\text{dor}}\):

\[
(225) \quad \text{zwaar-}\!_{\mathord{\text{or}}} \!\text{‘heavier’ (*zwaar-\(\mathord{\text{or}}\), boor-}\!_{\mathord{\text{or}}} \!\text{‘driller’ (*boor-\(\mathord{\text{or}}\))}\]
\]

An apparent exception is the suffix -\(\mathord{\text{on}}\) which serves many inflectional functions, like marking plurality on nouns and on verbs, and marking infinitival tense, and which can be added freely to any word ending in \(\mathord{\text{/n/}}\): *baan+\(\mathord{\text{on}}\) ‘jobs’, *wen+\(\mathord{\text{on}}\) ‘(we/to) adjust’. This may be related, however to the fact that \(\mathord{\text{/n/}}\) after schwa is often deleted.\(^{50}\)

\(^{48}\)A possible reason for this, suggested by van Oostendorp (2000), is that syllables which are phonetically headed by a schwa and closed by a sonorant, phonologically behave as if they have a syllabic sonorant in Dutch.

\(^{49}\)Or, to be more precise, by prefixation. There are forms which

\(^{50}\)The deletion takes place in large regions of the Dutch-speaking area, including most of the economically and culturally dominant ones. In the areas in which it does not take place, the suffix is usually realized as a syllabic \(\mathord{n}\).
The constraint in [224] may be reduced to the OCP: if we assume that schwa is a vowel without any features, the two vowels which neighbour to it on the lefthand and on the righthand side are adjacent to each other in a very obvious way.

Of course, the constraints just introduced run against the forms which might otherwise be coming out as the optimal form. They do not explain yet why those candidates are blocked from occurring. It will be our task in the next section to provide that piece of the puzzle.

7.4 Relativized MParse within Coloured Containment

We may conclude from the data in the previous two sections that diminutives in Standard Dutch pose yet another instance of the ineffability puzzle which forms the topic of this book. In the remainder of this chapter, I will first consider a solution which throws out these forms as unwanted; subsequently, this account will be embedded into a theory of language variation to account for the fact that the forms are not completely ungrammatical.

Let us reconsider the constraint MParse. We have seen in section [7.2] that Prince & Smolensky (1993) use this constraint in a fairly radical way: underlyingly there is no morphological structure at all, even the morphemes do not have an internal structure. Furthermore, it is assumed that the null parse has no morphological structure assigned to it at all.

Notice, however, that it is reasonable in its own right to have a less radical version of a constraint that says that morphological structure needs to be incorporated in the overall parse of the word, just like phonological structure. Indeed, it seems reasonable, given an (Items-and-Arrangement) theory of morphology within OT, to assume:

- That the input can be either an unstructured set of morphemes, or complex word consisting of morphemes arranged into some structure (given the morphological version of Richness of the Base).
- That the optimal output should consist of a morphological word (just like the optimal output will consist of a phonological word).
- That there are (M)PARSE constraints which require that individual morphemes should be part of the morphological structure.

The constraint MParse could be formulated as a family of constraints:

\[(226) \text{MParse}(M): \text{Every morpheme M has to be parsed into a morphological word.}\]

We assume that instances of this constraint exist for every individual morpheme. Since these are normal OT constraints, they can be violated, and we
expect them to be violated sometimes. In the case of the Dutch diminutive, consider for instance an input consisting of a preposition (aan ‘on’) plus a preposition. Dutch does not allow for diminution of determiners, probably for morphological reasons. We could thus set up the following analysis:

(227) \(\text{LEXDIM: There is no diminutive of function words.}\)

(228) 

<table>
<thead>
<tr>
<th></th>
<th>LEXDIM</th>
<th>MPARSE(DIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. aantje (+DIM)</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b. aan</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The winning form in (228a) does not parse the diminutive either morphologically or phonologically. If we have a more precise look at its phonological surface structure, we might observe the following:

\[
\begin{array}{c}
\mu \\
/ & \quad \text{DIM} \\
/ & \quad a : n \quad tj \quad \phi \\
/ & \quad \omega \\
\phi
\end{array}
\]

(229)

The diminutive suffix is parsed neither in the morphological structure nor in the phonological structure of the output. The reason why it is not parsed in the morphological structure is the constraint LEXDIM: there is no diminutive of the preposition aan. This also means that the winner candidate in (228) is not a diminutive, even though its input is (and the diminutive suffix is still hanging around in the output).

What about a candidate with the following structure?

\[
\begin{array}{c}
\mu \\
/ & \quad \text{DIM} \\
/ & \quad a : n \quad tj \quad \phi \\
/ & \quad \omega \\
\phi
\end{array}
\]

(230)
In this case, the segments connected to the diminutive show up in the phonology, but not in the morphology. This is clearly an undesirable result: we have a form which phonologically looks like a diminutive, although it is not a diminutive from a morphological (or semantic) point of view. But indeed it can be shown that forms such as these will never surface: a form such as (230) is harmonically bounded by a form such as (229).

The reasons for this are simple: both forms obviously satisfy all morphological constraints to the same extent; but they also both satisfy all the relevant phonological faithfulness constraints (PARSE-φ(α)). The reason for this is that those constraints require all morphologically affiliated material to be parsed into the phonology. Since the material belonging to the diminutive prefix is not part of the morphological structure, these constraints do not apply to them: the phonological parsing of this material in (230) thus is not necessary for reasons of phonological faithfulness.

On the other hand, the segments of the diminutive in (230) will probably violate markedness constraints of the type *STRUC, since they have material which is not present. Furthermore, they violate PARSE-μ(α), since they represent phonological material which is not incorporated into the morphological structure. In other words, these segments behave like epenthetic material in all relevant respects. Since epenthesis is not required in these cases — we know this because epenthetic material also does not show up in the ‘non-diminutive’ form of aan — we derive that (230) is indeed bounded by (229). The former will therefore never surface.

In this example, the diminutive is blocked by morphological well-formedness, but of course there is no a priori reason why phonological constraints could not similarly force morphological underparsing (which I have denoted in the following examples by putting the name of the relevant morpheme in parentheses). For instance, if we consider the forms ending in coronal obstruents, we get the following:

(231)

<table>
<thead>
<tr>
<th></th>
<th>FAITHFUL</th>
<th>OCP(cor)</th>
<th>MPARSE(DIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*lentø (DIM)</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>lentøtø</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lentøja</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Like in the case of the preposition just discussed, the relevant output form is not a diminutive in any relevant sense: it just is a non-diminutive form both with respect to its phonological shape as well with respect to its morphological, syntactic or semantic context. This is the sense in which it is not possible to form a diminutive of this word. Even if we set up the input in the most ideal way — with a diminutive morpheme — the output will ignore this input.
If we take into consideration the larger class of cases of words ending in schwa in (214), we only need to replace the relevant markedness constraint:

\[
\begin{array}{|c|c|c|}
\hline
\text{s} \ 	ext{schwa} /+ / \text{tj} @ & \text{FAITHFUL} & \*_{\text{o}_i \text{o}_j} \text{PARSE}(\text{DIM}) \\
\hline
\text{\textit{yav@}} & *! & \\
\text{\textit{ya:tj@}} & *! & \\
\hline
\end{array}
\]

Notice that this solution may be somewhat too strong. It throws out complete forms which speakers indeed consider to be strange or marked, but which nevertheless occur, and are also not necessarily considered to be totally ungrammatical. In other words, we need an analysis of the question mark, and a theory based on the MPARSE does not provide us with one directly (and neither does any other one). We will return to this in section 7.5.

The following example from Swedish (Iverson, 1981; Raffelsiefen, 2002; Féry & Fanselow, 2003; Rice, 2005) is a little more complicated. Adjectives get an ending \/-t/ in the neuter in attributive position:

(233) a. \textit{en rysk (MASC) pojke} ‘a Russian boy’
    b. \textit{et rysk-t (NEUTER) barn} ‘a Russian child’

However, if the adjectival stem ends in /d:/, the neuter form simply becomes impossible:

(234) a. \textit{en r"add (MASC) pojke} ‘a scared boy’
    b. *\textit{et r"add-t (NEUTER) barn} ‘a scared child’

In cases like this, we thus have a gap in the paradigm which is probably due to an OCP effect on the feature coronal. Furthermore, other (‘phonological’) options of resolving the OCP, for instance deleting the /t/ into one segment, are not available because of high-ranking faithfulness constraints. We then have the following tableau:

(235)\[
\begin{array}{|c|c|c|}
\hline
/\textit{r"add}/ & \text{OCP(cor)} & \text{PARSE(C)} & \text{MPARSE (NEUTER)} \\
\hline
\textit{\textit{r"add}}: (\text{NEUTER}) & * & \\
\textit{r"add}:t & *! & \\
\textit{r"add}: & *! & \\
\hline
\end{array}
\]

Again, the form we have thus created is not marked as bearing neuter agreement. This is why it has become unusable in (234), given a sufficiently high-ranking syntactic constraint enforcing agreement. A solution will then have
7.5. The question mark and language variation

We are left with one problem to address: why are some of the ‘unacceptable’ forms generated not completely unacceptable, but rather marked with a question mark? There is some literature on the question mark within the Optimality Theoretic framework, in works such as Anttila (1997, 2002); Boersma (1998, 2001); Boersma & Hayes (2001); Coetzee (2004). It is important to note that this literature is embedded in work on language variation: the grammar can generate more than one output, of which there are certain less preferred ones.

In the work of Anttila, which we will follow here, the grammar allows for some internal variation. In particular, constraints can be partially ordered. For instance, within the grammar of a language we might find three constraints A, B and C which are not ordered. This means that every ordering of these constraints (A ≫ B ≫ C, A ≫ C ≫ B) is equally likely. In the case of three constraints there are 3! = 6 possible orderings. Now suppose that four of these rankings give output α for some input a, while two give output β. This will then mean that β is in some sense more ‘marked’ than α as a pronunciation of a; β might get a question mark. With some modifications, irrelevant for our present discussion, similar considerations hold for the view on language variation defended in Boersma (1998, 2001); Boersma & Hayes (2001) and related works. Similarly, in Coetzee (2004), the output of the Evaluation is not a single candidate, but an ordered list, with the preferred candidate in the first position. The phonetics will usually pronounce the first element of the list, but in some cases, it may also pronounce an element with a lower position. These would then be perceived as having one or more question marks.

Combining one of these ideas with MPARSE may give us something similar to the effect we have observed for Dutch: the forms we find are question marked, since they are competing with a form which violates MPARSE. In the following, I choose Anttila’s approach for the sake of concreteness, but I suspect that the general idea outlined below could also be applied to the other frameworks. For instance, let us divide the constraint FAITHFUL in (232) into its relevant constituent parts — a constraint against deletion of schwa, PARSE-{o}, and a constraint against final devoicing, PARSE-[+voice], and let us assume that these two constraints are unranked with respect to
each other and with respect to MPARSE in the sense of Anttila. We then have the following six rankings (assuming that the constraint against sequences of schwas is ranked still on top, and that we therefore do not have to consider gavatja):

<table>
<thead>
<tr>
<th>/gav@+/tj@/</th>
<th>PARSE-∅</th>
<th>PARSE-[+voice]</th>
<th>MPARSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>gav@ (DIM)</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ya:tj@</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARSE-∅</td>
<td>MPARSE</td>
<td>PARSE-[+voice]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gav@ (DIM)</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>ya:tj@</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARSE-[+voice]</td>
<td>PARSE-∅</td>
<td>MPARSE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gav@ (DIM)</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>ya:tj@</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARSE-[+voice]</td>
<td>PARSE-∅</td>
<td>MPARSE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gav@ (DIM)</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>ya:tj@</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARSE-[+voice]</td>
<td>PARSE-∅</td>
<td>MPARSE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gav@ (DIM)</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>ya:tj@</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARSE-[+voice]</td>
<td>PARSE-∅</td>
<td>MPARSE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gav@ (DIM)</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>ya:tj@</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPARSE</td>
<td>PARSE-[+voice]</td>
<td>PARSE-∅</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The pronounced diminutive form thus sometimes wins, but most of the time it is bested by a form which does not function as a diminutive at all. This describes the ungrammatical feel that the diminutive form has.

An interesting aspect of Anttila (1997, 2002)'s work is that it relates the assignment of intuitive question marks by a native speaker to distribution of forms over a corpus. In this case, such a distribution can also be detected, for instance if we compare the diminutive forms of the names of the four seasons in the Google corpus.51

---

51 The Google-corpus is the corpus of Dutch texts as it is defined by the internet search engine Google.com. According to a fairly recent estimate (van Oostendorp, 2005a), the corpus contains 3,000,000,000 Dutch words. This means that the relative percentage of all the diminutives is fairly small. These searches were done on 28.III.2006. The results here have been gathered and normalized in the way described in van Oostendorp (2005a).
7.5. The question mark and language variation

<table>
<thead>
<tr>
<th>normal form</th>
<th>diminutive</th>
<th>plural of diminutive</th>
</tr>
</thead>
<tbody>
<tr>
<td>lente ‘spring’</td>
<td>lentetje 22</td>
<td>lentetjes 32</td>
</tr>
<tr>
<td>zomer ‘summer’</td>
<td>zomertje 779</td>
<td>zomertjes 236</td>
</tr>
<tr>
<td>herfst ‘autumn’</td>
<td>herfstje 108</td>
<td>herfstjes 89</td>
</tr>
<tr>
<td>winter ‘winter’</td>
<td>wintertje 894</td>
<td>wintertjes 696</td>
</tr>
</tbody>
</table>

There clearly is a difference between the word for ‘spring’ and the words for the other seasons. There is no obvious reason in lexical semantics why this could be the case, although a few remarks are in order. In the first place, there are big differences between the occurrences of names of different seasons, for which I have no explanation; notice however that, given the fact lente is a season name which occurs more often than zomer, for instance, we should expect there to be more instances of its diminutive rather than less, but this is not the case.

In the second place, the fact that the plural form of the diminutive of winter is that (De) Winter is a fairly frequent family name, and that the structure family name + diminutive + plural is a productive way of referring to members of a family:

(238) Dat de De Wintertjes niks uitvoeren in Amerika, baseer je op een mailtje van een zekere Emma Rose (wie is dat?).

You base the assumption that the family De Winter is not working at all in America, on an e-mail by a certain E.R. (who is that?)

Similar examples are harder to find for the other words, because they occur less frequently as family names. The following data are based on a database of Dutch familynames in 1993; the numbers indicate (roughly) the number of times the name was found in a Dutch register of family names on the internet:

(239) name | frequency  
--- | ---  
lente | 19  
zomer | 700  
herfst | 127  
winter | 1053  

The very low frequency of lentetjes may thus be partially ascribed to the infrequency of the family name Lente, but this is not true for singular lentetje, which has no such direct connection to a name.

52http://vetvetvet.web-log.nl/log/3858282
53The database has been built by Ann Marynissen of the University of Cologne and can be consulted at http://www.familienaam.be/.
54Notice that this result can be extended to all other theories which describe ineffability effects as the result of a constraint, which can interact with other constraints.
Finally, we have to turn to the observation that the forms can become better if they are repeated over time. I propose that the reason for this is that the forms become lexicalised; which means that the words become analysed as a whole. Suppose, for instance that the word *gavetje* becomes lexicalised.

We then obtain the following tableau:

<table>
<thead>
<tr>
<th></th>
<th>/gavetje/</th>
<th>FAITHFUL</th>
<th>*&lt;i&gt;_{\text{a}, \text{aj}}&lt;/i&gt;</th>
<th>MPARSE(DIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>/gavetje/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>/gavetje/</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>/gavetje/</td>
<td></td>
<td>!</td>
<td></td>
</tr>
</tbody>
</table>

In this case, the there is no relevant candidate with un ‘unparsed’ diminutive suffix, since there is no underlying diminutive suffix anymore. Therefore, ‘normal’ faithfulness decides that it is better to preserve the marked configuration.

Remember from the discussion on page 122 that the archaic suffix *ke* is more acceptable on names: most informants seem to think that *Hildotjo* is more acceptable than *Hildotja*. I suppose that this is also due to this process of lexicalisation. The fact that it is archaic means that it is not considered to be productive. I therefore suppose that it can only be interpreted as being input to the grammar as a whole.

<table>
<thead>
<tr>
<th></th>
<th>/hildotja/</th>
<th>FAITHFUL</th>
<th>*&lt;i&gt;_{\text{a}, \text{aj}}&lt;/i&gt;</th>
<th>MPARSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>/hildotja/</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>/hildotja/</td>
<td></td>
<td>!</td>
<td></td>
</tr>
</tbody>
</table>

One could wonder whether this analysis does not conflict with the principle of Richness of the Base. I do not think it does so in an essential way. Of course, it is possible in principle to posit two underlying morphemes /hilda/ and /ko/; the result of the evaluation of such a form would be ineffable, hence not pronounced, or question marked.

### 7.6 Conclusion

Representing ineffability seems to come at some cost within any branch of Optimality Theory; as far as I can see, no solution has been proposed which does not invoke some special mechanism that is somehow outside of the core of the OT system. This may be seen to imply that ineffability remains to be one of the real problems for the theory, maybe on a par with the issue of

---

55Strictly speaking, there is no issue of violating *<i>_{\text{a}, \text{aj}}</i>" anymore, since as we have seen, this constraint only applies to schwas of different morphological colour (belonging to different morphemes). But if a word is lexicalised, this means that all segments are in the same lexical item together, so they will all have the same morphological colour. But this means that potentially we can do away with the reference to colouring in [225].
opacity: if only additional patches can be used to describe a phenomenon which is as real as the unpronounceability of certain logically possible forms, the theory may be seen as incomplete at best.

I believe that the proposals put forward here, form a step forward in a few ways. First, they do not extend the computational power of the theory in any significant way. Second, it reduces the extra mechanism to a minimum — and even that minimum may be reduced to representational mechanisms which are independently necessary. But more importantly, we have shown how interaction with theories of language variation give us the possibility for describing the question mark. I have argued that this is an important advantage of MPARSE analyses over

The question arises to what extent other examples of ineffability discussed in the literature can also be argued to be question marks rather than asterisks. For instance, Féry & Fanselow (2003) argue that the German diminutive suffix gives an ineffable result when it is combined with a stem which ends in an unstressed full vowel:

\[
\begin{array}{ll}
\text{normal form} & \text{diminutive} \\
a. \text{Jahr} \ 'year' & \text{Jährchen} \\
b. \text{Brud} \ 'brother' & \text{Brüdorchchen} \\
c. \text{Monat} \ 'month' & *\text{Mönatchen}, *\text{Mönäthchen} \\
\end{array}
\]

Féry & Fanselow (2003) observe that the reason for this ineffability probably is the umlaut associated to the diminutive suffix in German: this wants to be linked to the stressed vowel in the stem, but apparently it gives a bad result if on its way to this stressed vowel it needs to either skip or pass through an unstressed full vowel.

Crucial here are the question marks. A Google search (on 28.III.2006) on the four forms for the diminutive gives the following:

\[
\begin{array}{|l|l|}
\hline
\text{diminutive} & \# \\
\hline
2\text{Monatchen} & 1,140 \\
2\text{Mönäthchen} & 154 \\
*\text{Mönatchen} & 217 \\
*\text{Mönäthchen} & 34 \\
\hline
\end{array}
\]

The judgements do not seem to correspond exactly to the Google findings; the reason for this may be that different dialects solve the problem in different ways. Importantly, however, we are obviously dealing also in this case also

---

56 Some of the findings in Google involve articles on the ineffability of the diminutive of Monat in German; but these are only 2 or 3, and do not differentiate between the different forms.

57 In some Dutch dialects, at least, umlaut to the stressed vowel is possible in words of this type, provided the unstressed vowel is reduced to schwa: foto 'picture' → fêtro (Gas-
with question marked forms rather than with forms which are completely ungrammatical.

Bibliography


Notice, by the way that this form has the properties of both German and Dutch ineffability but seems perfectly possible in the dialects in question, which are mostly spoken on the border between Germany and the Netherlands. More dialectological work into these dialects is clearly needed. In any case, the German orthographic form ‘Minatchen’ might represent a form which is pronounced with a reduced a as well.


Féry, Caroline (1994). ‘Umlaut and inflection in German’. ROA 33.


