WHAT IT MEANS TO BE A LOSER:

NON-OPTIMAL CANDIDATES IN OPTIMALITY THEORY

A Dissertation Presented

by

ANDRIES W. COETZEE

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2004

Department of Linguistics

© Copyright by Andries W. Coetzee 2004

All Rights Reserved

WHAT IT MEANS TO BE A LOSER:

NON-OPTIMAL CANDIDATES IN OPTIMALITY THEORY

A Dissertation Presented

by

ANDRIES W. COETZEE

Approved as to style and content by:

_____
John J. McCarthy, Chair

_____
John Kingston, Member

_____
Joseph V. Pater, Member

_____
Barbara H. Partee, Member

_____
Shelley Velleman, Member

_____
Elisabeth O. Selkirk, Department Head
Department of Linguistics

ACKNOWLEDGEMENTS

At the completion of this study there are many people that I need to thank. Foremost of these are John McCarthy. He has been a wonderful mentor and advisor. I also owe much to the other members of my committee – John Kingston, Joe Pater, Barbara Partee and Shelley Velleman. Being a member of the Linguistics community at UMass has been an academically enriching experience, and for that I owe thanks to every member of the faculty and to all my fellow students.

My time at UMass has been marked by personal tragedy and sadness. Through all the turmoil I have received constant support from the everyone involved in the Linguistics Department. I will always be deeply thankful to the Department for that.

I also want to acknowledge the role that my fellow classmates played in my life. The friendships that we formed made the graduate school experience easier and more fun. Minjoo Kim and Angela Carpenter deserve special mention. With them I could share my joys, fears and sadness. Thank you!

My parents have been a constant presence and support throughout. I also want to use this opportunity to thank them.

My wife, Lize, was always my most enthusiastic supporter, and I will forever love her for that. I wish that she could share with me the joys of completion.

Lastly, I want to thank all the people who are connected to 37 Grant Avenue – Alexia, Andrew, Stephanie and Marcy. You all have made my last semester at UMass the most memorable. And, of course, a very special thank you to David. No one has shared as much in the creation of this dissertation as you have, and for that I love you.

ABSTRACT

WHAT IT MEANS TO BE A LOSER:

NON-OPTIMAL CANDIDATES IN OPTIMALITY THEORY

SEPTEMBER 2004

ANDRIES W. COETZEE, B.A., POTCHEFSTROOM UNIVERSITY

M.A., POTCHEFSTROOM UNIVERSITY

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor John J. McCarthy

In this dissertation I propose a rank-ordering model of EVAL. This model differs from classic OT as follows: In classic OT, EVAL distinguishes the best candidate from the losers, but does not distinguish between different losers. I argue that EVAL imposes a harmonic rank-ordering on the complete candidate set, so that also the losers are ordered relative to each other. I show how this model of EVAL can account for non-categorical phenomena such as variation and phonological processing.

Variation. In variation there is more than one pronunciation for a single input. Grammar determines the possible variants and the relative frequency of the variants. I argue that EVAL imposes a harmonic rank-ordering on the entire candidate set, and that language users can access more than the best candidate from this rank-ordering. However, the accessibility of a candidate depends on its position in the rank-ordering. The higher a candidate appears, the more often it will be selected as output. The best candidate is then the most frequent variant, the second best candidate the second most

frequent variant, etc. I apply this model to vowel deletion in Latvian and Portuguese, and to [t, d]-deletion in English.

Phonological processing. Language users rely on grammar in word-likeness judgments and lexical decision tasks. The more well-formed a non-word, the more word-like language users will judge it to be. A more well-formed a non-word is considered more seriously as a possible word, and language users will be slower to reject it in a lexical decision task.

The rank-ordering model of EVAL accounts for this as follows: EVAL compares non-words and imposes a rank-ordering on them. The higher a non-word occurs in this rank-ordering, the more well-formed it is. Therefore, the higher a non-word occurs, the more word-like it will be judged to be, and the more slowly it will be rejected in lexical decision tasks.

I illustrate this by discussing two sets of experiments on how grammar influences phonological processing. The first set investigates the influence of the OCP on processing in Hebrew, and the second the influence of a constraint on [sCvC]-words in English.

CONTENTS

ACKNOWLEDGEMENTS ……... v

ABSTRACT ……... vi

CHAPTER

I. INTRODUCTION …….. 1

1. Theoretical preliminaries ……... 2

1.1 On being a better (or a worse) loser: a rank-ordering model of EVAL ……. 3
1.2 Extending the comparative powers of EVAL ……... 5

2. Applying the rank-ordering model of EVAL to phonological variation ……... 10

2.1 [t, d]-deletion in Jamaican English ……... 11
2.2 Accounting for the Jamaican pattern in a rank-ordering model of EVAL …….. 12

2.2.1 Intra-contextual variation ……. 12
2.2.2 Inter-contextual variation ……. 16
2.2.3 Limiting variation: the critical cut-off …….. 17
2.2.4 Summary of proposal …… 22

3. Applying the rank-ordering model of EVAL to the phonological processing of non-words …… 23

4. Structure of the dissertation ……... 27

II. A RANK-ORDERING MODEL OF EVAL …….. 29

1. A rank-ordering model of EVAL …….. 32
2. EVAL and the ordering associated with individual constraints ……. 36

2.1 Characterization of constraints …….. 36
2.2 Ordering the candidates with respect to individual constraints …… 40

2.2.1 Individual candidates in the set $\langle K/C, \leq_C \rangle$ …… 46
2.2.2 $\leq_C$ defines a chain ……. 52
2.2.3 The chain defined by $\leq_C$ always has a minimum …… 54
2.3.4 The set K/C is a partition on K ……. 56

3. EVAL and the ordering associated with the full constraint set …….. 59

3.1 Combining the ordered sets associated with individual constraints …… 60

3.1.1 Taking the Cartesian product …… 60
3.1.2 Imposing the lexicographic order on $K/C_\times$ ……... 63
3.1.3 Simplifying the set $K/C_\times$ and the ordering $\leq_\times$ ……... 69

3.2 The properties of K/C_{Com} and $\leq_{\text{Com}}$ ……... 75

3.2.1 $\langle K/C_{\text{Com}}, \leq_{\text{Com}} \rangle$ and the strictness of strict domination ……... 76
3.2.2 $\leq_{\text{Com}}$ defines a chain …….. 81
3.2.3 The chain defined by $\leq_{\text{Com}}$ always has a minimum …… 83
3.2.4 The set K/C_{Com} is a partition on K ……. 86
3.2.5 The members of K/C_{Com} as equivalence classes on K …….. 89

4. Summary …… 95

Appendix A: Definitions ……. 101
Appendix B: Theorems and Lemmas ……. 107

III. VOWEL DELETION IN LATVIAN ……. 110

1. The data ……. 111
2. The analysis ……... 112

2.1 Intra-contextual variation: more deletion than retention ……..... 113
2.2 Inter-contextual variation: more deletion in tri-syllabic than in bi-syllabic words ……. 118
2.3 Limits on variation …… 119
2.4 Ranking arguments ……... 128
2.5 Summary ……... 131

IV. VOWEL DELETION IN FAIALENSE PORTUGUESE …….. 133

1. The basic vocalic phonology of Faialense Portuguese …….. 135

1.1 The oral vowels of Faialense Portuguese …….. 135
1.2 Vowel reduction and deletion in unstressed syllables ……. 138

2. Vowel reduction in Faialense Portuguese ……. 141

2.1 The constraints …….. 142

2.1.1 Ranking conservatism ……... 147

2.2 Low vowels ……... 147
2.3 Back vowels …….. 149
2.4 Schwa ……..... 152
2.5 The high front vowel ……. 152
2.6 The mid front vowels …… 153
2.7 Summary ……... 155

3. The interaction of vowel reduction and deletion in Faialense Portuguese ……. 156

3.1 The data ……. 157
3.2 Variation between [ŭ] and ∅ …… 164

3.2.1 Non-final in a prosodic word: |ŭ ™ ∅| …….. 165
3.2.2 Final in a prosodic word: |∅# ™ ŭ#| …….. 169

3.3 Variation between [ə̆] and ∅ ……. 172

3.3.1 Non-final in a prosodic word: |ə̆ ™ ∅| ……... 173
3.3.2 Final in a prosodic word: |∅# ™ ə̆#| …….. 177

3.4 Interim summary ……... 179
3.5 Variation between [ĭ] and ∅ ……. 181
3.6 No variation: always [ĕ] …… 185
3.7 Final summary ……... 189

4. Variation across contexts ……... 191

4.1 Vowel quality …… 192

4.1.1 The non-undergoers …….. 193

4.1.1.1 The peak-affinity constraints in general ……... 198

4.1.2 The undergoers ……. 200

4.2 Position in the prosodic word …… 201
4.3 Stress of the following syllable ……. 203

4.3.1 The non-undergoers and *LAPSE …….. 204
4.3.2 The undergoers and *CLASH ……. 205
4.3.3 *LAPSE, *CLASH and the deletion of individual vowels …….. 208

4.4 Final summary ……... 209

5. Summary: vowel reduction and deletion in Faialense Portuguese ……. 211

V. [t, d]-DELETION IN ENGLISH …… 214

1. The following phonological context …….. 216

1.1 The data ……. 217
1.2 The analysis ……... 221

1.2.1 The constraints …….. 221
1.2.2 Accounting for the observed patterns ……... 228

1.2.2.1 Intra-contextual variation …….. 229
1.2.2.2 Inter-contextual variation …….. 237

1.2.3 Factorial typology – what are the possible dialects? …….. 242

1.2.3.1 Conditions for variation …… 244
1.2.3.2 Possible deletion patterns …….. 246
1.2.3.3 Impossible deletion patterns ……. 249

1.2.4 Two outstanding questions ……... 251

1.2.4.1 Only [t, d]? …… 252
1.2.4.2 What about [t, d] preceded by a vowel? …… 255

1.3 An alternative explanation …… 258

2. The preceding phonological context ……. 263

2.1 The data ……. 263
2.2 The analysis ……... 273

2.2.1 The constraints …….. 273
2.2.2 Accounting for the observed patterns ……... 278

2.2.2.1 Intra-contextual variation …….. 278
2.2.2.2 Inter-contextual variation …….. 282

2.2.3 Factorial typology – what are the possible dialects? …….. 284

2.2.3.1 Conditions for variation …… 285
2.2.3.2 Possible deletion patterns …….. 286
2.2.3.3 Impossible deletion patterns ……. 293

2.2.4 Two outstanding questions ……... 297

2.2.4.1 What about [p, b] and [k, g]? …… 297
2.2.4.2 What about [t, d] preceded by a vowel? …… 301

3. Considering alternatives …… 302

3.1 Conceptual problems ……. 306
3.2 Practical problems ……. 309

Appendix: Factorial typologies ……... 314

VI. WELL-FORMEDNESS JUDGMENTS AND LEXICAL DECISION …… 330

1. Preliminary considerations …… 332

1.1 Well-formedness and lexical decision in a rank-ordering model of EVAL …….. 333
1.2 Factors in the processing of non-words …… 336

1.2.1 Lexical statistics ……... 337
1.2.2 Grammar ……... 344

2. The OCP in the processing of non-words in Hebrew …… 348

2.1 Restrictions on identity in roots …… 349

2.1.1 Bi-consonantal roots ……. 349
2.1.2 From bi-consonantal roots to tri-consonantal stems ……. 353
2.1.3 The OCP: banning identical consonants ……... 359

2.1.3.1 No identical consonants in the surface realization of the root ……... 359
2.1.3.2 Avoiding identical consonants in the stem ……. 363

2.1.4 The critical cut-off …… 366
2.1.5 Summary ……... 368
2.1.6 Lexicon optimization …… 374
2.1.7 The processing of non-words ……... 377

2.2 Word-likeness and lexical decision in Hebrew ……. 380

2.2.1 Well-formedness judgment experiments …….. 381

2.2.1.1 Gradient well-formedness judgment experiments ……... 381
2.2.1.2 Comparative well-formedness judgment experiments ……... 385

2.2.2 Lexical decision experiments ……... 388

3. The processing of non-words of the form [sCvC] in English ……... 393

3.1 Restrictions on [sCvC]-words ……... 395

3.1.1 A basic OT analysis …….. 398

3.1.1.1 The critical cut-off …… 401
3.1.1.2 Interim summary ……... 402

3.1.2 Motivation of analysis …….. 402

3.1.2.1 The place harmony scale in English …….. 403
3.1.2.2 Constraints against the co-occurrence of homorganic voiceless stops …….. 406
3.1.2.3 On the markedness of [s+stop]-structures ……. 409
3.1.2.4 *sCvC-constraints as local conjunctions …….. 417

3.1.3 The processing of non-words of the form [sCvC] …….. 420

3.2 Word-likeness and lexical decision in English ……. 423

3.2.1 Well-formedness judgment experiments …….. 426

3.2.1.1 Gradient well-formedness judgment experiment …… 426
3.2.1.2 Comparative well-formedness judgment experiment …… 433

3.2.2 Lexical decision experiment ……. 440

3.2.2.1 Experimental design …….. 441
3.2.2.2 Results ……... 443

4. Considering alternatives …… 446

4.1 Grammar or lexical statistics? ……... 447

4.1.1 Hebrew and the OCP …… 447
4.1.2 [sCvC]-forms in English ……... 451

4.2 Alternative grammatical analyses ……. 453

4.2.1 Classic OT …… 454
4.2.2 Stochastic OT ……... 456
4.2.3 Crucially unranked constraints ……. 464

5. Concluding remarks ……... 468

Appendix: Tokens used in [sCvC]-experiments ……. 471

BIBLIOGRAPHY …….. 475

Winning isn't everything. It's the only thing.

Vince Lombardi

In determining the correct – optimal – parse of an input, as the constraint hierarchy is descended, each constraint acts to disqualify competitors with absolute independence from all other constraints. A parse found wanting on one constraint has absolutely no hope of redeeming itself by faring well on any or all lower-ranking constraints.

Prince and Smolensky (1993:78)

CHAPTER 1

INTRODUCTION

In this dissertation I propose a new way to think about how EVAL, the evaluative component of an Optimality Theoretic grammar, works. In particular, I propose two changes in the way that EVAL is usually viewed in classic OT (Prince and Smolensky, 1993): (i) First, I propose a change in the information structure that EVAL imposes on the candidate set. In classic OT, EVAL is seen as defining only a two-level distinction in the candidate set – i.e. EVAL distinguishes between the best candidate and the set of losers, but does not impose any structure on the set of losers. I argue that EVAL imposes a harmonic rank-ordering on the complete candidate set – i.e. also the losers are ordered from best to worst. Rather than just a two-level ordering, EVAL imposes a multi-level harmonic ordering on the candidate set. I therefore refer to the model of EVAL that I propose as the "rank-ordering model of EVAL". (ii) Secondly, I argue that we should expand the comparative powers of EVAL. In classic OT EVAL is assumed to compare candidate sets of only one kind, namely sets generated by GEN as candidate outputs for a specific input. (I will refer to these sets as "generated comparison sets".) Candidates in a generated comparison set are all morphologically related to each other via a shared input. However, I will show that EVAL can compare any set of candidates, even candidates that are not related to each other via a shared input (i.e. EVAL can also compare with each other mappings such as the following: /in_1/ $\rightarrow$ [$cand_1$] and /in_2/ $\rightarrow$ [$cand_2$], where /in_1/ and /in_2/ are morphologically unrelated).

The motivation for this alternative view of EVAL is two-fold. First, I will show that this alternative view of EVAL does not require any changes to the architecture of EVAL in a classic OT grammar. Even in a classic OT grammar EVAL generates the information necessary to impose a harmonic rank-ordering on the complete candidate set, and even in classic OT EVAL has the ability to evaluate non-generated comparison sets. I am not really proposing a change to the architecture of a classic OT grammar. I am rather pointing out two previously unnoticed and/or unappreciated features of an OT grammar. Secondly, I will show that the different conceptualization of EVAL extends the empirical coverage of an OT grammar. Specifically, it enables us to account for non-categorical phenomena. I will discuss two kinds of non-categorical phenomena in detail in this dissertation, namely variation in production (more than one grammatical pronunciation for the same word) and the processing of non-words (well-formedness judgments and lexical decision).

The rest of this introductory chapter is structured as follows. In §1 I will discuss the alternative view of EVAL that I am proposing in more detail. In §2 I will then present a brief illustration of how this alternative view of EVAL can be used to account for variation in production, and in §3 I will show how it can be used to account for the processing of non-words. Finally, in §4 I will explain the structure of the rest of the dissertation.

1. Theoretical preliminaries

I am proposing two innovations to the way that we standardly think about an OT grammar. Neither of the innovations requires a change to the architecture of the grammar.

One of the innovations makes use of information that a classic OT grammar generates but that is usually considered irrelevant in OT literature, namely about the relationships between the non-optimal candidates in the candidate set. The other entails an extension of the comparative powers of EVAL so that it can compare candidates from different, morphologically unrelated inputs. Comparing such morphologically unrelated forms is something that even the EVAL of classic OT could do. However, this ability of EVAL was never appreciated in classic OT. Each of these two extensions is discussed in more detail below.

1.1 On being a better (or a worse) loser: a rank-ordering model of EVAL

OT is a theory of winners. EVAL makes only one distinction in the candidate set, namely between the winning candidate and the mass of losers. There is no claim made about the relationships between the losers. There is no concept such as being a better or a worse loser.

This standard view of OT is held in spite of the fact that EVAL has the power to make finer grained distinctions in the candidate set. If we were to remove the optimal candidate from a candidate set, and submit only the set of losers to EVAL, then EVAL will again identify the one loser that is better than all the rest. This best loser can then be removed, and we can allow EVAL to compare only the smaller set of remaining losers, again identifying the best one from this smaller set. In fact, this process can be repeated for as long as there are still candidates left, and we can therefore rank-order the full candidate set in this way.

In classic OT the assumption is that EVAL imposes only a two-level harmonic ordering on the candidate set (the optimal candidate against the rest). Alternatively, we

can also entertain the possibility that EVAL imposes a harmonic rank-ordering on the full candidate set. These two views about the information structure that EVAL imposes on the candidate set can be represented graphically as in (1). Candidates appearing higher are more harmonic relative to the constraint ranking.

(1)

Standard OT view: 2 levels	**Alternative view: Rank-Ordering**
$\{Can_x\}$	$\{Can_x\}$
|	|
$\{Can_y, Can_z, Can_w, \ldots\}$	$\{Can_y\}$
	|
	$\{Can_z\}$
	|
	$\{Can_w\}$
	|
	…

In classic OT the losers are lumped together in an amorphous set. It does not matter whether you are the second best candidate or the worst candidate – all that counts is that you are not the winner. In the rank-ordering model of EVAL this is not the case. Even the losers are ordered with respect to each other. It is now possible to be a better or a worse loser – i.e. being the second best candidate is now something that has meaning.

Under the standard assumption that every constraint evaluates every candidate, the information necessary to rank-order the full candidate set is automatically generated by an OT grammar. It is not necessary to make any additions to the way that an OT grammar works in order to get this information. The assumption in classic OT that EVAL imposes only a two-level ordering on the candidate set is therefore not a theoretical necessity. It is rather the case that this information is available, but that classic OT assumes it to be irrelevant. If this information were really irrelevant, then an OT grammar

would have been much too powerful, generating a mass of irrelevant information. I will therefore argue for the opposite, namely that this information is relevant and that it available to and accessed by language users in non-categorical phenomena. This lends strong support to the general architecture of an OT grammar.

Since the alternative view about the output of the grammar is that EVAL imposes a rank-ordering on the full candidate set, I will refer to this model as the "rank-ordering model of EVAL".

1.2 Extending the comparative powers of EVAL

The basic function of EVAL is to compare a set of candidates in terms of their relative harmony. In classic OT it is assumed that the set that EVAL compares is the set of candidate output forms for a single input – i.e. the candidates generated by GEN for some input. However, there is nothing in the way EVAL functions that requires the comparison set to be of this kind. EVAL is blind to the origin of the candidate set that it compares, and in principle EVAL can compare any set of candidates, whether they all come from the same input or each from a different input.

To make this discussion more concrete, consider a language *L* that tolerates codas, i.e. with the ranking ||{MAX, DEP} o NOCODA||. Suppose that *L* had only two lexical entries, namely /pata/ and /kispa/. Since *L* tolerates codas, both of these underlying representations should be mapped faithfully onto the surface. This is easily confirmed by allowing EVAL to evaluate the candidate sets generated by GEN for each of /pata/ and /kispa/. This is the kind of comparison that is traditionally done in OT. In (2) I show the tableaux for these comparisons.

(2) **EVAL evaluates generated comparison sets**

a. /pata/ → [pa.ta]

/pata/	MAX	DEP	NOCODA
pa.ta			
pa.tak		*!	*
pat	*!		*

b. /kispa/ → [kis.pa]

/kispa/	MAX	DEP	NOCODA
kis.pa			*
ki.si.pa		*!	
ki.pa	*!		

In (2a) the faithful candidate does not violate any constraint, and is therefore necessarily selected as output. The tableau in (2b) confirms that *L* will not avoid codas by either epenthesis or deletion.

But this is not the only informative comparison that we can make. Even though *L* tolerates codas (as shown in (2b)), it is still the case that a form with a coda is more marked than a form without a coda – i.e. although both [pa.ta] and [kis.pa] are possible words of *L*, [kis.pa] is more marked because it earns a violation of NOCODA while [pa.ta] does not. This intuition can be expressed formally by allowing EVAL to compare these two candidates. This is shown in the tableau in (3).

(3) **Comparison between the faithful candidates**

		MAX	DEP	NOCODA
2	/kispa/ → [kis.pa]			*
1	/pata/ → [pa.ta]			

Two things can be remarked about the tableau in (3). First, although these two candidates are not morphologically related to each other via a shared input, they can be compared in a straightforward manner. No additions have to be made to the way in which EVAL works in order to compare two such morphologically unrelated candidates. Secondly, this tableau supplies us with information that we would not have access to if we considered only the generated comparison sets of classic OT. That a possible word with a coda is somehow less well-formed than a possible word without a coda is not expressed in the tableaux in (2). These tableaux show only that words with and without codas are allowed in *L*. The difference between words with and without codas can only be expressed in comparisons such as that done in tableau (3). Of the two forms compared, the form without the coda is rated better by EVAL. (In this tableau I indicate this with Arabic numerals – the 1 next to the second candidate indicates that it is rated best, the 2 next to the first candidate that it is rated second best. More on these conventions follows in §2.)

EVAL therefore has the power to compare candidates that are not related to each other via a shared input, and comparisons like these can provide us with information that is not otherwise available. I will show in the rest of this dissertation how information from such comparisons is used in accounting for non-categorical phenomena.

What are all the comparison sets that EVAL can evaluate? In principle there is no limit on the sets of candidates that EVAL can compare. If we collect together in one set the candidates that GEN generates for all inputs, EVAL can compare any subset of this large set. In (4) I formalize what the sets are that EVAL can evaluate.

(4) **Comparison sets**

Let *Input* = $\{in_1, in_2, \ldots in_n\}$ be the set of all possible inputs, and GEN(in_i) the set of candidates generated by GEN for input in_i.

Then the set of all possible candidates is the arbitrary union of the sets generated by GEN for all possible inputs, i.e. $\cup_i$ GEN(in_i) for all in_i in *Input*.

EVAL can compare any subset of candidates from $\cup_i$ GEN(in_i). The set of possible comparison sets is therefore the powerset of this set, i.e. $\wp(\cup_i \text{GEN}(in_i))$.

The set generated by GEN for some single input in_j is a special kind of comparison set. It is namely that member of $\wp(\cup_i \text{GEN}(in_i))$ that includes all and only the members of GEN(in_j). In order to distinguish this sub-class of possible comparison sets from the rest of the possible comparison sets, I will refer to these special comparison sets as *generated comparison sets* – to emphasize that these are the sets generated by GEN for specific inputs. Comparison sets that do not qualify as generated comparison sets will simply be called *non-generated comparison sets*.

Consider again language *L* from above. Recall that we assumed that *L* has only two lexical forms, /pata/ and /kispa/. For *L* the set *Input* is then {/pata/, /kispa/}. The set of all possible candidates in *L* is the set (GEN(/pata/) $\cup$ GEN(/kispa/)), and EVAL can compare any subset of (GEN(/pata/) $\cup$ GEN(/kispa/)), i.e. any member of the powerset $\wp$(GEN(/pata/) $\cup$ GEN(/kispa/)). In (5) I list a few of comparison sets for *L* that EVAL can evaluate.

(5) **Comparison sets in *L***

a. **Generated comparison sets**

GEN(/pata/) = {/pata/ → [pa.ta],
/pata/ → [pa.ta.ka],
/pata/ → [pat], …}

GEN(/kispa/) = {/kispa/ → [kis.pa],
/kispa/ → [ki.si.pa,
/kispa/ → [ki.pa], …}

b. **All possible candidates**

GEN(/pata/)
∪ GEN(/kispa/) = {/pata/ → [pa.ta],
/kispa/ → [kis.pa],
/kispa/ → [ki.si.pa],
/pata/ → [pat], …}

c. **Non-generated comparison sets**

Set_1 = {/pata/ → [pa.ta], /kispa/ → [kis.pa]}

Set_2 = {/pata/ → [pat], /kispa/ → [ki.pa]}

Set_3 = {/pata/ → [pa.tak], /kispa/ → [ki.si.pa]}

Etc.

EVAL can compare any of the sets from (5a) or (5c). The sets in (5a) are the generated comparison sets of classic OT. In (2) above I showed how EVAL would compare the candidates from these two sets. The first of the non-generated comparison sets in (5c) consists of the two faithful candidates from GEN(/pata/) and GEN(/kispa/). This set was evaluated by EVAL in (3) above.

2. Applying the rank-ordering model of EVAL to phonological variation

In this dissertation I will argue that EVAL imposes a harmonic rank-ordering on the full candidate set – as shown in (1) above. The output of the grammar therefore contains more information than just what the best candidate is. It also contains a wealth of information about how the non-optimal candidates (the losers of classic OT) are related to each other. I will argue that this enriched information is available to language users, and that they access it *inter alia* in non-categorical phenomena. In this section I will provide a brief discussion of how this enriched information can be used to explain variation in production.

There is a large body of literature on phonological variation, both within sociolinguistics and within formal linguistic theory (e.g. Anttila, 1997, 2002, Anttila and Cho, 1998, Anttila and Revithiadou, 2000, Anttila and Fong, 2002, Guy, 1980, 1981, 1991a, 1991b, 1994, 1997, Guy and Boberg, 1997, Guy and Boyd, 1990, Hinskens, 1996, Hinskens *et al*., 1997, Kiparsky, 1993, Nagy and Reynolds, 1997, Patrick, 1991, Reynolds, 1994, Reynolds and Sheffer, 1994, Silva, 1988, 1991, 1993, 1997, 1998). One thing that stands out from all of this literature is an acknowledgement that phonological variation is at least partially grammatically determined. If there is more than one pronunciation possible for a specific word, then grammar plays a role in determining how likely each variant is to be observed. In order to account for this aspect of phonological variation, it is therefore necessary to rely on grammar.

I will argue that the rank-ordering model of EVAL can be used to account for the role that grammar plays in phonological variation. In this section I will give a brief illustration of how this can be done. In this discussion I will use variable [t, d]-deletion in

Jamaican English as an example. The purpose of this discussion is to explain the theoretical assumptions that I make, and not to provide a full account of [t, d]-deletion in Jamaican English. For a detailed discussion of this phenomenon, see Chapter 5 below. The rest of this section is structured as follows: In §2.1, I give a small sample of the data on [t, d]-deletion in Jamaican English. In §2.2, I then show how the rank-ordering model of EVAL can be used to account for these data.

2.1 [t, d]-deletion in Jamaican English

In English, a [t, d] that appears as last member of a word-final consonant cluster is subject to variable deletion – i.e. a word such as *west* can be pronounced as either [wɛst] or [wɛs]. This phenomenon has been studied extensively over the past four decades and I discuss it in detail in Chapter 5. Here I will discuss only one aspect of this process in Jamaican English as reported by Patrick (1991).

Patrick reports that the likelihood of [t, d]-deletion depends *inter alia* on what follows on the [t, d]. [t, d] followed by a consonant is more likely to delete than [t, d] followed by a vowel. The table in (6) contains the relevant data (Patrick, 1991:181).

(6) **[t, d]-deletion in Jamaican English**

	Pre-C	Pre-V
n	1,252	793
% deleted	85	63

When a word like *west* is followed by a consonant, as in *west bank*, there is an 85% percent probability that it will be pronounced without the final [t] as [wɛs bæŋk], and a 15% probability that it will be pronounced with the final [t] as [wɛst bæŋk]. However, when it is followed by a vowel, as in *west end*, there is a 63% that the final [t]

will not be pronounced as in [wɛs ɛnd], and a 37% chance that the final [t] will be pronounced as in [wɛst ɛnd].

In this phenomenon we are dealing with two variants, the retention candidate and the deletion candidate. There are two aspects of this variation pattern that need to be accounted for. (i) *Intra-contextual variation*. For any given input, we have to account for the fact that one of the two variants occur more frequently than the other. In both pre-vocalic and pre-consonantal context, the deletion variant occurs more frequently in Jamaican English. (ii) *Inter-contextual variation.* Although deletion is preferred over retention both pre-vocalically and pre-consonantally, it is still the case that pre-consonantal context shows higher deletion rates than pre-vocalic context. In §2.2 I explain how each of these two aspects of the variation phenomenon is accounted for in the rank-ordering model of EVAL.

2.2 Accounting for the Jamaican pattern in a rank-ordering model of EVAL

2.2.1 Intra-contextual variation

For any given input, EVAL evaluates the generated comparison set and imposes a harmonic rank-ordering on this set. Language users can potentially access all the candidates in this rank-ordering. However, the accessibility of a candidate depends directly on how high it occurs in the rank-ordering. The candidate that occupies the highest slot in the rank-ordering is most accessible and is most likely to be selected as output. This will therefore be the most frequently observed variant. The candidate that occupies the second slot in the rank-ordering, is second most accessible, and will the second most frequent variant, etc.

In both pre-consonantal and pre-vocalic context the deletion candidate is the most frequently observed candidate. (I will use ∅ to stand for the deletion candidate in the discussion below.) EVAL therefore has to rate the deletion candidate better than the retention candidate in both of these contexts, i.e. EVAL needs to impose the following rank-ordering on these two candidates |∅ ™ t/d|. This can be achieved only if the highest ranked constraint that distinguishes between the deletion and the retention candidate is a constraint that favors deletion over retention.[1] The deletion candidate obviously violates the anti-deletion constraint MAX. I will assume the existence of the markedness constraints in (7a) that militate against the retention of [t, d] in pre-vocalic and pre-consonantal position. I also assume the fixed ranking in (7b) between these two constraints. For motivation of these constraints and this ranking, see Chapter 5 §1.2.1.

(7) a. **Markedness constraints**

***Ct#C**: A word-final [t, d] is not allowed if it is both preceded and followed by a consonant.

***Ct#V**: A word-final [t, d] is not allowed if it is preceded by another consonant and followed by a vowel.

b. **Ranking**

||*Ct#C o *Ct#V||

[1] The idea of constraints favoring or disfavoring a candidate comes from Samek-Lodovici and Prince (1999). Let $C(x)$ represent the number of violations constraint C assigns to candidate x, and K the set of all candidates under consideration. For some candidate *Can* to be favored by constraint C, the following statement must then be true: $\neg\exists\ k \in K\ (C(k) < C(Can))$. Conversely, for some candidate *Can* to be disfavored by constraint C, the following statement must be true: $\exists\ k \in K\ (C(k) < C(Can))$.

Since we need the deletion candidate to be preferred over the retention candidate in both contexts, we need the markedness constraints to outrank MAX, i.e. ||*Ct#C o *Ct#V o MAX||. The tableaux in (8) show how EVAL will evaluate the generated comparison sets for a pre-vocalic and pre-consonantal input with this ranking.

(8) **Generated comparison sets evaluated**

a. **Pre-consonantal context**

/…Ct # C/	*Ct#C	*Ct#V	MAX
1 ∅			*
2 t	*		

Output of EVAL

∅ MAX
|
t *Ct#C

b. **Pre-vocalic context**

/…Ct # V/	*Ct#C	*Ct#V	MAX
1 ∅			*
2 t		*	

Output of EVAL

∅ MAX
|
t *Ct#V

In (8a) we see that the deletion candidate violates only MAX while the retention candidate violates only *Ct#C. Because MAX is the lower ranked of these two constraints, the deletion candidate is rated better by EVAL. This is indicated in two ways. First, by the Arabic numerals next to the candidates. The numeral 1 next to the deletion candidate indicates that this is the candidate rated best by EVAL – i.e. this is the candidate that

occupies the highest slot in the rank-ordering that EVAL imposes on the candidate set, and also the optimal candidate of classic OT. The numeral 2 next to the retention candidate indicates that it is rated second best by EVAL. Below the tableau I also give a graphic representation of the rank-ordering that EVAL imposes on the candidate set. Candidates that appear higher on this ordering are rated better by EVAL. Every candidate is also indexed with the highest ranked constraint that disfavors the candidate. I also indicate each of the candidates that are observed as outputs by the familiar pointing hand in this graphic representation.

The claim that I make in this dissertation is that language users have access to the full candidate set via the rank-ordering that EVAL imposes on the candidate set. Unlike in classic OT, the language user can therefore also access candidates other than the best or optimal candidate. This explains why both the best candidate (deletion) and the second best candidate (retention) are observed as outputs. Even though the retention candidate is not the best candidate, it is still accessible to language users via the rank-ordering. However, the accessibility of a candidate depends on the position it occupies in this rank-ordering – the higher position it occupies the more accessible it is. Since the deletion candidate occupies a higher slot than the retention candidate in both (8a) and (8b), the deletion candidate in both pre-vocalic and pre-consonantal context is more accessible than the retention candidate. The prediction is therefore that the deletion candidate will be observed as output more frequently than the retention candidate in both of these contexts.

Variation is possible because language users can also access non-optimal candidates. The relative frequency of different variants is accounted for by the fact that not all candidates are equally accessible. There is one complication here – in (8) I

consider only two candidates. The generated comparison set for any input obviously contains many more than just two candidates. In §2.2.3 below I will come back to these other candidates. In §2.2.2 I first show how to account for the inter-contextual variation.

2.2.2 Inter-contextual variation

In the previous section I have shown how the rank-ordering model of EVAL can be used to account for the variation in the pronunciation of individual inputs. However, this is not the only relevant aspect of variation. We can now explain why deletion is preferred over retention in both pre-consonantal and pre-vocalic context. But deletion is preferred even more in pre-consonantal than pre-vocalic context – pre-consonantal context has a deletion rate of 85% while pre-vocalic context has a deletion rate of only 63%. We also need to account for this. In order to account for inter-contextual variation, I use the ability of EVAL to evaluate non-generated comparison sets.

The driving force behind the deletion of final [t, d] is to avoid violation of the markedness constraints *Ct#C and *Ct#V. Violation of these two constraints can be avoided by deletion, i.e. at the expense of a MAX-violation. The retention candidate in pre-consonantal position violates *Ct#C, while the retention candidate in pre-vocalic position violates *Ct#V. Because of the ranking ||*Ct#C o *Ct#V|| retention in pre-consonantal position is more marked than retention in pre-vocalic context. The drive to delete in pre-consonantal context is therefore stronger than in pre-vocalic context. This explains why pre-consonantal context has a higher deletion rate. This intuition can be captured formally by allowing EVAL to compare the faithful retention candidates from pre-vocalic and pre-consonantal contexts. This comparison is shown in (9).

(9) **Non-generated comparison set: the faithful candidates**

		*Ct#C	*Ct#V	MAX
2	/…Ct # t/ → [Ct # t]	*		
1	/…Ct # V/ → [Ct # V]		*	

Output of EVAL

Pre-vocalic: /…Ct # V/ → [Ct # V] $_{*Ct\#V}$

Pre-consonantal: /…Ct # C/ → [Ct # C] $_{*Ct\#C}$

The comparison in (9) shows that retention in pre-consonantal context is more marked than retention in pre-vocalic context. A greater decrease in markedness can therefore be bought by a MAX-violation (by deletion) in pre-consonantal than in pre-vocalic context. This explains why more deletion is observed in pre-consonantal than pre-vocalic context. By allowing EVAL to evaluate non-generated comparison sets, we can also explain why different contexts show different variation patterns.

2.2.3 Limiting variation: the critical cut-off

In the discussion of the intra-contextual variation §2.2.1 I considered only two candidates for each input. However, the generated comparison set for each input contains many more candidates than just these two. Under the rank-ordering model of EVAL, each of these candidates will occupy a slot in the rank-ordering. And under the assumption that the language user has access to levels below the highest level in this ordering, we predict that these other candidates should also be accessed as variants, even if less frequently. This problem becomes particularly acute when we consider candidates other than the deletion candidate that also avoid violation of the constraints *Ct#C and *Ct#V. In addition to deletion of the final [t, d], there are many other ways in which violation of

these constraints can be avoided. For instance, it is possible to insert a vowel between the [t, d] and the preceding consonant. A phrase such as *west bank* can therefore be pronounced with deletion of the final [t] as [wɛs bæŋk], or with insertion of a vowel between [s] and [t] as [wɛsət bæŋk]. Both of these pronunciations avoid violating *Ct#C, both will be in the generated comparison set of /wɛst bæŋk/, and both will occupy a slot in the rank-ordering that EVAL imposes on this comparison set. Why is only the deletion candidate ever observed as a variant pronunciation for *west bank*?

Jamaican English tolerates deletion but not epenthesis in order to avoid violation of *Ct#C. Stated in terms constraint violation: Jamaican English is willing violate MAX but not DEP in order to avoid a violation of *Ct#C. Even though every candidate in the generated comparison set occupies a slot in the rank-ordering that EVAL imposes on this set, and even though the whole candidate set is in principle accessible to the language user via this rank-ordering, the language user will not access this rank-ordering to an arbitrary depth. There are certain constraints that a language is willing to violate – in the case at hand here those constraints are the markedness constraints *Ct#C and *Ct#V, and the faithfulness constraint MAX. But there are also constraints that a language is not willing to violate unless if absolutely required. From the discussion above, it follows that Jamaican English is not willing to violate DEP. I propose that there is a critical cut-off on the constraint hierarchy that divides the constraint set into those constraints that a language is willing to violate and those that a language is not willing to violate. A candidate disfavored by a constraint ranked higher than the cut-off will not be accessed as output if there is a candidate (or candidates) available that is not disfavored by any constraint ranked higher than the cut-off.

In Jamaican English both the deletion and the retention candidate are accessed as variant outputs. Both of these candidates are therefore disfavored only by constraints ranked lower than the critical cut-off. This means that MAX, *Ct#C and *Ct#V rank lower than the critical cut-off. The epenthetic candidate is never accessed as a variant output, and this candidate should be disfavored by at a constraint ranked higher than the cut-off. The epenthetic candidate violates the anti-epenthesis constraint DEP, and this constraint therefore ranks higher than the cut-off. In the tableau in (10) I reconsider the generated comparison sets from (8) above. However, this time I include the epenthetic candidate, the constraint DEP, and the critical cut-off. The critical cut-off is indicated by a thick vertical line in the tableaux. Constraints to the left of this line are ranked higher than the cut-off, and constraints to the right of this line are ranked lower than the cut-off. More discussion of the conventions used in these tableaux follows below the tableaux.

(10) **The generated comparison sets with epenthetic candidates**

Ranking: ||DEP o Cut-off o*Ct#C o*Ct#V o MAX||

a. **Pre-consonantal context**

/..Ct # C/	DEP	*Ct#C	*Ct#V	MAX
1 C#C				*
2 Ct#C		*		
3 CVt#C	*!			

Output of EVAL

Deletion: C#C $_{\text{MAX}}$

Retention: Ct#C $_{\text{*Ct#C}}$

Cut-off

Epenthesis: CVt#C $_{\text{DEP}}$

((10) continued)

b. **Pre-vocalic context**

/..Ct # V/	DEP	*Ct#C	*Ct#V	MAX
1 C#V				*
2 Ct#V			*	
3 CVt#V	*!			

Output of EVAL

Deletion: C#V $_{\text{MAX}}$

Retention: Ct#V $_{\text{*Ct\#V}}$

Cut-off

Epenthesis: CVt#V $_{\text{DEP}}$

The epenthetic candidate violates DEP, which is ranked higher than the critical cut-off. Since there are candidates available that are not disfavored by any constrains ranked higher than the cut-off, the epenthetic candidate will not be accessed as a variant output. This is indicated in two ways in these tableaux. I use the exclamation mark to mark the DEP-violation of the epenthetic candidate. This exclamation mark has the same meaning as in classic OT tableaux – it indicates the violation that is responsible for eliminating a candidate as a potential output. On the graphic representation of the rank-ordering that EVAL imposes on the candidate set, I draw a solid horizontal line indicating the position of the critical cut-off. Candidates appearing higher than this line are not disfavored by any constraints ranked higher than the cut-off. Candidates appearing below this line are disfavored by at least one constraint ranked higher than the cut-off. Since there are candidates appearing above this line, the language user will not access any candidates below this line. The pointing hands marking the deletion and

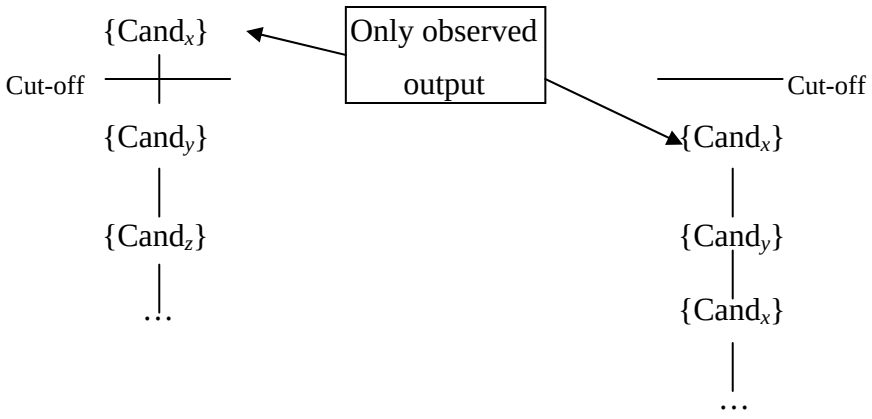
retention candidates indicate that they are both accessed as outputs. The epenthetic candidate does not have a pointing hand, indicating that it is not a possible output. The epenthetic candidate stands in here for all candidates other than the deletion and retention candidate – that is, all other candidates are also disfavored by at least one constraint ranked higher than the cut-off.

Although all candidates appear in the rank-ordering that EVAL imposes on the candidate set and although the language user has potential access to the complete rank-ordered candidate set, he/she will not normally access this rank-ordering to an arbitrary depth. In general, candidates disfavored by constraints ranked higher than the cut-off are not accessed.

Given this account of variation, how can categorical phenomena be accounted for? what conditions must be met for language users to access only the best candidate in the rank-ordering that EVAL imposes on the candidate set? The discussion in the rest of this section is general in nature and does not pertain specifically to the Jamaican data above. There are two ways in which a categorical phenomenon can arise: (i) If all but one candidate are disfavored by at least one constraint ranked higher than the cut-off, then the single candidate that is not disfavored by a constraint ranked higher than the cut-off will be the only observed output. This is exactly like a variable phenomenon, except that there is only one variant that is observed 100% of the time. (ii) It is also possible that all candidates are disfavored by at least one constraint ranked higher than the cut-off. When this happens the language user has no option but to access candidates that are disfavored by a constraint ranked higher than the cut-off. In order to minimize the number of such candidates that appear as outputs, only the single best candidate is accessed in such a

situation. These two ways in which a categorical phenomenon can be modeled are represented graphically in (11).

(11) **Categorical phenomena in a rank-ordering model of EVAL**



2.2.4 Summary of proposal

I propose that variation should be explained in the following way: (i) EVAL imposes a harmonic ordering on the complete generated comparison set for every input, and the language user can potentially access all the candidates on this harmonic ordering. However, the accessibility of a candidate is directly related to the position it occupies in the rank ordering. The lower the position a candidate occupies in the rank-ordering, the less accessible it is. This explains why certain variants occur more often than others. The most frequent variant is the candidate rated best by the grammar, the second most frequent variant is the candidate rated second best, etc.

(ii) EVAL can compare morphologically unrelated forms (non-generated comparison sets). This property of EVAL is responsible for explaining why a variable process applies more or less frequently in different contexts. Suppose that the relevant non-generated comparison set contains the fully faithful candidates from two contexts,

and that the fully faithful candidate of one context is more marked than the fully faithful candidate of another context. Then the drive to be unfaithful is stronger in the more marked context, and this context is then predicted to be subject to higher rates of application of the variable process.

(iii) There is a critical cut-off point on the constraint hierarchy of every language. The language user will not access a candidate that is disfavored by a constraint ranked higher than the critical cut-off point if a candidate is available that is not disfavored by any such constraint. This explains why variation is limited to a few variants per input, and also why there are some contexts in which no variation is observed.

3. Applying the rank-ordering model of EVAL to the phonological processing of non-words

In this dissertation I will argue that the rank-ordering model of EVAL can also be used to account for the role that grammar plays in the phonological processing of non-words. Specifically, it can explain (i) how grammar influences judgments on the well-formedness of non-words, and (ii) how grammar influences the reaction times in lexical decision tasks. There is a large body of literature on well-formedness judgments and lexical decision tasks (e.g. Balota and Chumbley, 1984, Berent and Shimron, 1997, Berent *et al*., 2001a, 2001b, 2002, Frisch *et al*., 2001, Frisch and Zawaydeh, 2001, Hayes, 1997, 1998, Hayes and MacEachern, 1996, 1997, Pierrehumbert *et al*., In press, Shulman and Davison, 1977, Stone and Van Orden, 1993, Vitevitch and Luce, 1999). Although much of this literature focuses more on the non-grammatical aspects of phonological processing, there is still a general acknowledgement that grammar does play some part in phonological processing. In particular, two generalizations that arise from this literature

are that: (i) a non-word that is more well-formed according the grammar of some language is also judged to be more well-formed by speakers of that language, (ii) the less well-formed a non-word is according to grammar of some language, the more quickly speakers of that language are in detecting it as a non-word in a lexical decision task.

In this section I will illustrate briefly how the rank-ordering model of EVAL can be used to account for these generalizations. I will use data from well-formedness judgments and lexical decision experiments that I conducted with speakers of English. The purpose of the discussion here is not to give a full account of these data, but rather to illustrate how the rank-ordering model of EVAL can be used to account for this kind of data. A detailed discussion of the data can be found in Chapter 6 §3.

English allows words of the form [sTvT], but not of the form [sKvK] or [sPvP] – i.e. *state* is a word but **spape* and **skake* are not even possible words (Browne, 1981, Clements and Keyser, 1983, Davis, 1982, 1984, 1988a, 1988b, 1989, 1991, Fudge, 1969, Lamontagne, 1993: Chapter 6). I argue that there is a constraint against each of these types of forms – i.e. *sTvT, *sKvK and *sPvP. In English, *sKvK and *sPvP outranks some faithfulness constraint so that forms that would violate these constraints will never surface faithfully. However, *sTvT ranks lower than all relevant faithfulness constraints, so that forms that violate this constraint will surface faithfully. This therefore gives us the ranking ||{*sPvP, *sKvK} o Faithfulness o *sTvT||. On the grounds of cross-linguistic data and other phonotactic restrictions in English, I argue that there is also a ranking between *sPvP and *sKvK, namely ||*sPvP o *sKvK||. The complete ranking is therefore ||*sPvP o *sKvK o Faithfulness o *sTvT||.

I conducted an experiment in which I presented subjects with pairs of non-words of the form (i) [sTvT]~[sKvK], (ii) [sTvT]~[sPvP], and (iii) [sKvK]~[sPvP]. The task of the subjects was to select from each pair the non-word that is most well-formed. In the [sTvT]~[sKvK]-pairs subjects chose [sTvT] more often (75% of time), in the [sTvT]~[sPvP]-pairs they also chose [sTvT] more (78% of the time), and in the [sKvK]~[sPvP]-pairs they selected [sKvK] more (55% of the time). In general, the response pattern can therefore be summarized as: [sTvT] > [sKvK] > [sPvP]. This corresponds exactly to the assumed ranking between the three *sCvC constraints – non-words that violate the lowest ranked *sTvT constraint are chosen most frequently, then non-words that violate the second highest ranked constraint *sKvK, while non-words violating the highest ranked *sPvP are chosen least frequently.

I also conducted a lexical decision experiment. In this experiment I presented English speakers with a list of words and non-words, and their task was to decide for each token whether it is a word or a non-word. The non-words included [sTvT]-, [sKvK]- and [sPvP]-forms. I recorded the reaction times on correct non-word responses for these three kinds of non-words. The finding was that [sTvT] non-words were detected most slowly (mean response time = 403.53 ms), [sKvK] non-words more quickly (mean response time = 350.45 ms), and [sPvP] non-words the most quickly (mean response time = 303.33 ms). This also corresponds to the ranking between the *sCvC-constraints. The higher ranked the markedness constraint that a non-word violates, the more quickly it is identified as a non-word.

These response patterns can be explained by allowing EVAL to compare non-words of the form [sTvT], [sKvK] and [sPvP]. Such a comparison is shown in (12).

(12) **Comparing [sTvT], [sKvK] and [sPvP]**

	*sPvP	*sKvK	Faithfulness	*sTvT
sTvT				*
sKvK		*		
sPvP	*			

Output of EVAL

[sTvT] *sTvT
|
[sKvK] *sKvK
|
[sPvP] *sPvP

The higher a form appears in the rank-ordering that EVAL imposes on the comparison set, the more well-formed it is judged to be by language users. Similarly, the higher a form appears on this rank-ordering, the more seriously it is considered as a possible word, and the longer it takes to identify it as a non-word. The relation between this rank-ordering, and well-formedness judgments and lexical decision reaction times is represented graphically in (13).

(13) **Rank-ordering model of EVAL, well-formedness judgments and lexical decision reaction times**

Well-formedness judgments	**Rank-ordering imposed by EVAL**	**Lexical decision RT's**
Decreasing well-formedness judgments ↓	[sTvT]	↓ Decreasing RT in lexical decision
	|	
	[sKvK]	
	|	
	[sPvP]	

Recall that I am proposing two changes in the way that we think about EVAL: (i) First, I assume that EVAL imposes a harmonic rank-ordering on the full comparison set, and does not only distinguish the best candidate from the rest. (ii) Secondly, I assume that EVAL can evaluate candidates that are not morphologically related (non-generated comparison sets). Both of these assumptions are used in explaining the [sCvC]-data.

First, the three forms that are being compared in the tableau in (12) are not morphologically related. EVAL is therefore evaluating a non-generated comparison set here. Had EVAL only been able to compare candidate sets generated by GEN for some input, this comparison would not have been possible at all. Secondly, the output of EVAL in (12) consists of a three-level ordering. EVAL does more than distinguish between the best [sTvT] candidate and the other two candidates. EVAL also orders the two non-best candidates relative to each other. Since the subjects in the experiment treated [sKvK]- and [sPvP]-forms differently, we need a grammar that can distinguish between them. EVAL therefore generates information even about how the non-best (losing) candidates are related to each other, and this information is available to and accessed by language users.

4. Structure of the dissertation

This dissertation contains five content chapters. Chapter 2 is theory oriented. In this chapter I develop a set theoretic model of EVAL. This chapter is somewhat independent from the rest of the dissertation. It contains many general results about an OT grammar of which only some are directly relevant to this dissertation. The results that are relevant are: (i) I show that the rank-ordering model of EVAL is completely compatible with standard

assumptions about an OT grammar. The rank-ordering model of EVAL does not require any changes or additions to the architecture of an OT grammar. (ii) I also show that nothing in the way that EVAL works depends on the origin of the set of candidates that is being compared. Allowing EVAL to compare non-generated candidate sets therefore does not require any additions to the architecture of an OT grammar. This chapter will not be easily accessible to the non-mathematically inclined reader. I would suggest that these readers skip Chapter 2 and start reading at Chapter 3.

The rest of the dissertation illustrates empirical applications of the rank-ordering model of EVAL. In Chapters 3 to 5 I illustrate the application of this model to phonological variation. I discuss three examples: (i) Chapter 3 deals with variable vowel deletion in Latvian, (ii) Chapter 4 deals with variable vowel deletion in Faialense Portuguese, and (ii) Chapter 5 deals variable deletion of word-final coronal stops in English. Chapter 3 is a short chapter and serves an introductory purpose. In this chapter I explain the theoretical assumptions that I make in more detail. I suggest that readers start with this chapter. However, Chapters 4 and 5 are independent from each other and can be read separately.

In Chapter 6 I apply the rank-ordering model of EVAL to the phonological processing of non-words. I discuss two examples: (i) How the OCP influences well-formedness judgments and lexical decision in Hebrew, and (ii) how *sCvC-constraints influence well-formedness judgments and lexical decision in English.

CHAPTER 2

A RANK-ORDERING MODEL OF EVAL

An Optimality Theoretic grammar (Prince and Smolensky, 1993) consists of three major components, namely a generative function (GEN), a universal set of constraints (CON), and an evaluative function (EVAL). GEN is a universal function that generates a candidate set for any given input. CON contains all of the constraints that make up Universal Grammar. EVAL uses the constraints in CON to compare the candidates generated by GEN with each other in order to determine the output associated with an input. EVAL can therefore be considered as the center of an OT grammar – this is where, given the candidate set and the constraints, the output of the grammar is determined.

In this chapter of the dissertation I investigate the formal properties of EVAL in more detail. Specifically, I develop a set theoretic model of EVAL. This chapter serves two purposes: (i) In the rest of this dissertation I will argue that EVAL works somewhat differently from what has traditionally been assumed in the OT literature. I will make two claims about EVAL. First, rather than just distinguishing between the best candidate and the mass of losers, EVAL imposes a harmonic rank-ordering on the full candidate set. Secondly, EVAL can compare any set of candidates, irrespective of whether they are related to each via a shared input or not. (For more on these two claims, refer to Chapter 1.) This chapter shows that this alternative view of EVAL is entirely compatible with the architecture of a classic OT grammar and does not require any formal changes to the architecture of the grammar. (ii) But this chapter also serves the more general purpose of providing a mathematical model of EVAL. Once EVAL is formulated as a mathematical

object, many of the assumptions about an OT grammar that are implicitly part of OT literature, can be explicitly stated and formally proved.

The basic approach in this chapter is that of "explication" which Carnap defines as "transforming a given more or less inexact concept into an exact one, or rather, replacing the first by the second" (Carnap, 1962:3). Carnap calls the inexact concept that explication sets out to formalize the "explicandum", and the result of the explication process the "explicatum". The goal of this chapter is therefore not to propose a new theory of grammar, but rather to express in explicit, formal terms what is generally accepted about OT. Since explication is basically a descriptive activity, it is quite hard to decide whether the explicatum is "correct" or not (Carnap, 1962:4).[1] The correctness of the explicatum should be measured by how well it resembles the explicandum. However, since the explicandum is by definition not an exact, precise concept, it is difficult (if not in principle impossible) to determine whether the explicatum exactly fits the explicandum. Even so, we should be confident that the explicatum at least agrees with our basic intuitions about the explicandum. Throughout the discussion, I will therefore point out how the model of EVAL that I am developing resembles what is standardly accepted about an OT grammar. One of the most characteristic features of an OT grammar is the so-called "strictness of strict domination" principle (McCarthy, 2002b:4, Prince and Smolensky, 1993:78, 1997:1604). I will therefore in particular show that the model of EVAL developed here abides by this principle (see §3.2.1).

---

[1] See also Kornai (1995:xix-xxi) for a discussion of this same problem specifically with regard to formalizing linguistic theories.

The chapter is structured as follows: In §1 I give a short review of how the rank-ordering model of EVAL is different from what is usually assumed about EVAL in OT literature. I also discuss previous mathematical formalizations of EVAL, and show that they are not consistent with a rank-ordering model of EVAL. Section §2 gives a characterization of constraints as functions from the candidate set into ù, and then shows how the candidate set can be ordered with respect to individual constraints. In §3 I show how the orderings associated with individual constraints are combined into one single ordering for the whole grammar. The chapter ends in two appendices. Appendix A contains a list of all the definitions used in the chapter, and Appendix B a list of the theorems and lemmas formulated.

This chapter is somewhat independent from the rest of the dissertation. Since it contains many results about an OT grammar that are not directly relevant to the rest of the dissertation, it can be read as a self-contained unit without reading the rest of the dissertation. Similarly, the rest of the dissertation can also be read without reading this chapter. Any of the issues discussed in this chapter that are relevant elsewhere in the dissertation, are also discussed where they are relevant. However, this chapter contains the only comprehensive formal treatment of the theoretical assumptions made in this dissertation. Reading this chapter will enhance the overall understanding of the theoretical claims. Readers who are not mathematically inclined may either skip this chapter completely, or jump ahead in this chapter to §4. In section §4 I provide a brief summary the chapter, and point out which of the results of this chapter will be relevant in the rest of the dissertation.

1. A rank-ordering model of EVAL

Classic OT is a theory of winners. It makes only one distinction in the candidate set, between the winning candidate and the mass of losers. Once a candidate has been eliminated from the race, it is demoted to the set of non-optimals or losers. And once in this set of losers, all information supplied about the candidate by the constraints becomes irrelevant. All losers are treated alike – as members of one large amorphous set.

This standard view of an OT grammar is held in spite of the fact that the theory can make finer grained distinctions in the candidate set. If we remove the optimal candidate from a candidate set and consider only the set of losers, then there will again be a candidate that is better than all the rest. This best candidate amongst the losers can then be removed, and we can repeat the comparison again to find the best candidate in the remaining smaller set of losers. In fact, this process can be repeated for as long as there are still candidates left, and we can therefore rank-order the full candidate set in this way.

These two views about the output of an OT grammar are represented graphically in (1). Candidates appearing higher are more harmonic relative to the constraint ranking. The "alternative view" is the view that I am assuming in this dissertation.

(1)

Standard OT view	**Alternative view**
$\{Can_x\}$	$\{Can_x\}$
|	|
$\{Can_y, Can_z, Can_w, \ldots\}$	$\{Can_y\}$
	|
	$\{Can_z\}$
	|
	$\{Can_w\}$
	|
	…

Since the information about the relationships between losers is considered irrelevant in classic OT, previous mathematical models of EVAL were formulated to ignore this information. These models can be divided into two groups: (i) models that are formulated such the information about the relationships between the non-winners is not generated at all; (ii) models that generate this information but ignore it. I will discuss one example of each kind here.

As an example of the first kind, consider the model developed by Samek-Lodovici and Prince (1999, Prince, 2002). They view constraints as functions from sets of candidates to sets of candidates. A constraint takes as input a set of candidates and returns as output the subset consisting of those candidates that perform best on the particular constraint. All other candidates are demoted to the set of losers, irrespective of how they are related to each other. The set of losers might contain candidates that violate the particular constraint only once and also candidates that violate the constraint many times. No distinction is made between losers. EVAL is then simply a composition of the different constraint functions in the order in which they are ranked. For instance, if K is the candidate set generated by GEN, and $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$ the constraint hierarchy, then EVAL will have the form $(C_n \circ \ldots \circ C_2 \circ C_1)(K) = C_n(\ldots C_2(C_1(K)))$. EVAL then returns only the single best (set of) candidate(s) as output. The complement in K of the set returned by EVAL is the set of losers. But no information is available about how the candidates in this set of losers are related to each other. Other models that were developed along similar lines are those of Eisner (1997a, 1997b, 1999), Hammond (1997) and Karttunen (1998).

Moreton (1999) has a different conceptualization of EVAL, and his model serves as an example of the second type of model of EVAL. In Moreton's model the information about the relationships between losers is generated but ignored. The end result is that his model still makes only a two-level distinction in the candidate set. For Moreton, constraints are functions from the set of candidates into ù. A constraint takes a single candidate as input, and then maps the candidate onto the natural number corresponding to the number of times that the candidate violates the particular constraint. The fact that constraints are functions on the candidate set implies that every constraint applies to every candidate. This is the crucial difference between Moreton's model and the models discussed above. In the other models, every constraint prunes the candidate set down so that later constraints may not get the opportunity of applying to the full candidate set. Since all constraints evaluate all candidates in Moreton's model, the losers can in principle also be compared with each other. However, as shown below, Moreton conceptualizes EVAL in such a way that this does not happen.

Moreton defines for each candidate a *score vector*. The score vector of a candidate consists of the number of violations afforded the candidate by each of the constraints, ordered according to the ranking between the constraints. For instance, consider a constraint hierarchy $||C_1 \circ C_2 \circ C_3||$, and a candidate k that violates C_1 once, C_2 three times, and C_3 twice, i.e. $C_1(k) = 1$, $C_2(k) = 3$, and $C_3(k) = 2$. The score vector associated with k is then $v_k = \langle C_1(k), C_2(k), C_3(k)\rangle = \langle 1, 3, 2\rangle$. Every candidate has such a score vector associated with it. Score vectors are compared as stated in (2).

(2) **Comparing score vectors in Moreton's model**

Let $v = \langle s_1, s_2, \ldots, s_n \rangle$ and $v' = \langle s'_1, s'_2, \ldots, s'_n \rangle$ be score vectors.

We say that $v < v'$ iff $\exists\, j \leq n$ such that:

(i) $\forall\, i < j$: $s_i = s'_i$, and

(ii) $s_j < s'_j$

The score vector of some candidate k_1 precedes the score vector of some other candidate k_2 if the highest ranked constraint that judges the two candidates differently favors k_1 over k_2. Moreton then defines the output of the grammar (of EVAL) for some input as that candidate whose score vector precedes the score vectors of all other candidates. Although the information about the relationships between the other candidates is generated, this information is ignored in the final output of the grammar where only the best candidate is distinguished from the mass of losers. De Lacy has a similar characterization of constraints (de Lacy, 2002:30).

These earlier models are therefore not compatible with a rank-ordering model of EVAL. Models like that of Samek-Lodovici and Prince (1999) are in principle incompatible since they do not even generate the information that would be required to rank-order the full candidate set. Models such as those of Moreton (1999) generate this information and are therefore in principle compatible with a rank-ordering model of EVAL. However, these models are formulated in such a way that this information is ignored.

In the rest of this chapter I will develop a model of EVAL that generates the information about the relationships between the losers, and also uses this information

explicitly to impose a rank-ordering on the full candidate set. Although Moreton's model can in principle be extended to do this, I choose to formulate a model that is different from Moreton's. In Moreton's model, comparison between candidates is done in terms of score vectors. In the model that I develop comparison is done in terms of individual constraints – every constraint imposes a rank-ordering on the candidate set, and the orderings associated with different constraints are then combined to yield a final ordering for the full grammar. EVAL orders the candidate set in two stages, first in terms of individual constraints and then by combining the orderings associated with individual constraints. The next two sections discuss each of these two stages in the ordering process.

2. EVAL and the ordering associated with individual constraints

The comparison that EVAL makes is based on the violations that the constraints assign to each candidate. Before we can consider the properties of EVAL, it is therefore necessary to have a clear idea of what constraints are. In this section I will first characterize constraints, and only then show how EVAL orders the candidate set with respect to individual constraints.

2.1 Characterization of constraints

A constraint considers each of the candidates generated by GEN separately, and evaluates that candidate according to some substantive requirement. A constraint assigns a violation mark to a candidate for every instance of non-compliance of the candidate with the specific substantive requirement of that constraint. A constraint therefore sets up a

relation between a candidate and a number of violations. We can regard the domain of this relation as the set of all candidates, and its range as a subset of ù, the natural numbers. Constraints can then be characterized as in (3).

(3) **Constraints as relations between the candidate set and ù**

Let CON be the universal set of constraints, and K the set of candidates to be evaluated. Then, $\forall C \in$ CON:

C: $K \rightarrow$ ù such that $\forall\ k \in K$, $C(k) =$ number of violations of k in terms of C

This characterization of constraints is basically the same as that assumed by Moreton (1999).[2] However, it is significantly different from the view taken by Samek-Lodovici and Prince (Prince, 2002, Samek-Lodovici and Prince, 1999). For them, constraints take as argument not individual candidates, but sets of candidates. Also, a constraint does not return a natural number as its value, but a subset of the candidate set that it took as argument. (See discussion above in §1.)

---

[2] It is also in principle identical to the way that de Lacy (2002:30) defines constraints. For de Lacy a constraint is a relation that takes as input a candidate, and returns not a natural number but a set of violation marks. Comparison between candidates is then done by comparing the cardinality of the sets of violation marks of each candidate. However, since a repeated identical element in a set does not change the set (i.e. {*} = {*, *, *, …}), de Lacy has to introduce a method to make multiple violation marks distinct. He needs a way in which a set with *n* violation marks will have a cardinality of *n*. This is his solution: "To avoid this problem, take a 'violation mark' to be any element from a denumerably infinite set of discrete elements (e.g. the natural numbers). Thus, a set of three violation marks is {1,2,3}, with a cardinality of 3." (2002:30). A constraint is then a relation that maps each candidate onto a set with cardinality equal to the number of times that the candidate violates the constraint.

This is in principle identical to the characterization of constraints that I give in (3) above. The natural numbers can be reconstructed in set theoretic terms such that each natural number is simply a set with cardinality equal to the specific natural number, i.e. the natural number *n* is a set with cardinality *n* (Enderton, 1977:66-89). When we think about the natural numbers in set theoretic terms, then the way in which constraints are characterized above in (3) can be seen as relations that map each candidate onto a set with cardinality equal to the number of times that the candidate violates the constraint. De Lacy's characterization of constraints is therefore only superficially different from the view that I take.

Based on some generally accepted properties of an OT grammar we can show that constraints are functions. The definition of a function in (4) comes from Partee *et al*. (1993:30).

(4) **<u>Def. 1</u>: Functions**

A relation *R* from *A* to *B* is a function iff:

(a) the domain of *R* = *A* (i.e. every member of *A* is mapped onto some member of *B*), and

(b) each element in *A* is mapped onto just one element in *B* (*R* is single valued).

(5) **<u>Theorem 1</u>: Constraints as functions**

All constraints are functions.

It is not possible to prove Theorem 1. The truth of this Theorem does not follow from some inherent property of what it means to be a constraint, but from the way in which constraints are conventionally formulated in OT. The discussion in this paragraph is therefore not a proof, but only illustrative in nature. In order for (4a) to hold of constraints, it is necessary that every constraint assign some value to every candidate. This follows from the assumption that a constraint applies even to candidates that do not violate the constraint – it assigns the natural number zero to these candidates. In order for (4b) to hold, a constraint should assign a unique value to each candidate. This is obviously true. A candidate violates a constraint a fixed number of times, and this number of violations is the only value that the constraint can assign to a candidate.

Throughout this chapter I will use an example to illustrate the concepts that I discuss. In this example I will assume a grammar with only three constraints, i.e. CON = $\{C_1, C_2, C_3\}$ and only five candidates, i.e. $K = \{c_1, c_2, c_3, c_4, c_5\}$. I will also assume a specific ranking between the constraints, namely $\|C_1 \circ C_2 \circ C_3\|$. The tableau in (6) shows how each of the five candidates is evaluated by the three constraints in this example.

(6) **$\{c_1, c_2, c_3, c_4, c_5\}$ evaluated by $\|C_1 \circ C_2 \circ C_3\|$**

	C_1	C_2	C_3
c_1	*	**	
c_2	**	*	*
c_3		*	**
c_4	*	**	
c_5	**		**

In (3) constraints were characterized as relations between K and ù. This can now be illustrated: Since c_1 violates C_1 once, it means that C_1 will map c_1 onto the value 1, i.e. $C_1(c_1) = 1$. Similarly, $C_1(c_2) = 2$, $C_1(c_3) = 0$, etc. We can do the same for all three constraints and all five candidates. We can also represent each constraint as a set of ordered pairs $\langle x, y\rangle$ where x is a candidate and y the value onto which the constraint maps y. In (7) I show these sets of ordered pairs for every constraint.

(7) **Constraints as relations between *K* and ù**

$C_1 = \{\langle c_1, 1\rangle, \langle c_2, 2\rangle, \langle c_3, 0\rangle, \langle c_4, 1\rangle, \langle c_5, 2\rangle\}$

$C_2 = \{\langle c_1, 2\rangle, \langle c_2, 1\rangle, \langle c_3, 1\rangle, \langle c_4, 2\rangle, \langle c_5, 0\rangle\}$

$C_3 = \{\langle c_1, 0\rangle, \langle c_2, 1\rangle, \langle c_3, 2\rangle, \langle c_4, 0\rangle, \langle c_5, 2\rangle\}$

Theorem 1 (5) stated that constraints are functions. It is clear that the relations in (7) are functions. First, each of the five candidates in *K* is represented by an ordered pair in each of these sets. Secondly, every candidate is mapped onto only one value.

2.2 Ordering the candidates with respect to individual constraints

A constraint sees candidates only in terms of their violations. Two candidates that earn the same number of violations in terms of some constraint are therefore indistinguishable from each other as far as that constraint is concerned.[3] This means that the two candidates [a.ta] and [pu.i.ma], although they are clearly distinct, cannot be distinguished in terms of the constraint ONSET (every syllable must have an onset). Both of these candidates contain one onsetless syllable, and they are therefore both mapped onto the same value by the constraint ONSET, i.e. ONSET([a.ta]) = ONSET([pu.i.ma]) = 1. Although these two candidates are distinct, they share with each other all ordering relationships in terms of the constraint ONSET to the rest of the candidate set.

When we consider the ordering that EVAL imposes on the candidate set with reference to a specific constraint, it is therefore not necessary to consider an ordering that refers to every candidate individually. Rather, the ordering can be viewed as an ordering defined on sets of candidates, specifically on sets of candidates that share the same number of violations. This leads to a significant simplification by reducing the number of discrete elements that need to be compared.

This can also be illustrated with the example introduced from (6) above. In this example candidates c_2 and c_5 are clearly distinct – since they violate C_2 and C_3 to

[3] See Samek-Lodovici and Prince (1999) about such "grammatically indistinct" candidates. Sections §2.2.1 and §3.2.5 below contain more discussion of grammatical indistinctness in OT.

different degrees. Even so, in terms of C_1 they are indistinguishable – they both violate C_1 twice. Candidates c_2 and c_5 will therefore occupy the same slot in the ordering that EVAL imposes on the candidate set in terms of C_1. The same is true of candidates c_1 and c_4. In the same manner we can also establish the ordering that each of C_2 and C_3 will impose on the candidate set. In (8) I represent the orderings associated with each of these constraints graphically. Candidates (more accurately, sets of candidates) that appear higher in this graphic representation are rated better by the particular constraint.

(8) **Orderings imposed by EVAL on candidate set in terms of C_1, C_2 and C_3**

C_1	C_2	C_3
$\{c_3\}$	$\{c_5\}$	$\{c_1, c_4\}$
|	|	|
$\{c_1, c_4\}$	$\{c_2, c_3\}$	$\{c_2\}$
|	|	|
$\{c_2, c_5\}$	$\{c_1, c_4\}$	$\{c_3, c_5\}$

Since this ordering is clearly on sets of candidates and not on candidates, the first thing we need to do is to gather individual candidates into the sets on which the ordering is defined. The candidates in each of these sets are those candidates with the same number of violations in terms of the specific constraint. We can therefore define a relation that will express what the candidates in each of these sets have in common.

(9) **<u>Def. 2</u>: The relation $\approx_C$ on K**

Let K be the candidate set to be evaluated by EVAL, and CON the set of constraints.

Then, for all $k_1, k_2 \in K$, and for all $C \in$ CON, let:

$$k_1 \approx_C k_2 \quad \text{iff} \quad C(k_1) = C(k_2).$$

Consider C_1 in the example discussed above. Candidates c_1 and c_4 both earn 1 violation in terms of C_1, i.e. $C_1(c_1) = C_1(c_4) = 1$. From this it follows that c_1 and c_4 stand in the $\approx_{C1}$–relation to each other, or $c_1 \approx_{C1} c_4$. Since = is reflexive, we of course also have $c_4 \approx_{C1} c_1$. And since = is also symmetric we also have $c_i \approx_{C1} c_i$ for each candidate c_i. In the same way the $\approx_{Ci}$-relation can be determined for each constraint. In (10) I show these relationships for each of the constraints.

(10) **$\approx_C$-relations for each constraint**

$C_1(c_i) = 0$: $c_3 \approx_{C1} c_3$

$C_1(c_i) = 1$: $c_1 \approx_{C1} c_4$, $c_4 \approx_{C1} c_1$, $c_1 \approx_{C1} c_1$, $c_4 \approx_{C1} c_4$

$C_1(c_i) = 2$: $c_2 \approx_{C1} c_5$, $c_5 \approx_{C1} c_2$, $c_2 \approx_{C1} c_2$, $c_5 \approx_{C1} c_5$

$C_2(c_i) = 0$: $c_5 \approx_{C2} c_5$

$C_2(c_i) = 1$: $c_2 \approx_{C2} c_3$, $c_3 \approx_{C2} c_2$, $c_2 \approx_{C2} c_2$, $c_3 \approx_{C2} c_3$

$C_2(c_i) = 2$: $c_1 \approx_{C2} c_4$, $c_4 \approx_{C2} c_1$, $c_1 \approx_{C2} c_1$, $c_4 \approx_{C2} c_4$

$C_3(c_i) = 0$: $c_1 \approx_{C3} c_4$, $c_4 \approx_{C3} c_1$, $c_1 \approx_{C3} c_1$, $c_4 \approx_{C3} c_4$

$C_3(c_i) = 1$: $c_2 \approx_{C3} c_2$,

$C_3(c_i) = 2$: $c_3 \approx_{C3} c_5$, $c_5 \approx_{C3} c_3$, $c_3 \approx_{C3} c_3$, $c_5 \approx_{C3} c_5$

We can show that the relation $\approx_C$ is an equivalence relation, which basically means that two elements that stand in the $\approx_C$-relation to each other are indistinguishable from each other in terms of this relation. In (11) I state the requirements that must be met for a relation to be an equivalence relation (Enderton, 1977:56), and in (12) I then show that $\approx_C$ is indeed an equivalence relation.

(11) **<u>Def. 3</u>: An equivalence relation**

A binary relation R on some set is an equivalence relation on that set iff R is (i) reflexive, (ii) symmetric, and (iii) transitive.

(12) **<u>Theorem 2</u>: $\approx_C$ as an equivalence relation**

For all $C \in$ CON, $\approx_C$ is an equivalence relation on K.

This is obviously true. $\approx_C$ is by definition a binary relation. Also, $\approx_C$ is defined in terms of the relation = on ù, and = is reflexive, symmetric and transitive on ù. The relation $\approx_C$ therefore inherits these properties from =.

Since $\approx_C$ is an equivalence relation, we can use $\approx_C$ to define equivalence classes (Enderton, 1977:57) on the candidate set. An equivalence class in terms some equivalence relation R is a set containing all the forms that stand in the relation R to each other. In terms of the relation $\approx_C$ there will therefore be an equivalence class containing the candidates that earn zero violations in terms of C, a class containing the candidates that earn one violation in terms of C, a class containing the candidates that earn two violations in terms of C, etc.

(13) **<u>Def. 4</u>: Equivalence classes on K in terms of $\approx_C$**

For all $k_1 \in K$, $\{k_{1,,C} := \{k_2 \in K \mid k_1 \approx_C k_2\}$

Every candidate k_1 will therefore be grouped together into a set (an equivalence class) with all the other candidates that receive the same number of violations as k_1 in

terms of C. In (14) I show the equivalence classes that are defined on K in terms of each of the $\approx_C$-relations in our example grammar from (6).

(14) **Equivalence classes on the candidate set in terms of $\approx_C$-relations**

$C_1(c_i) = 0$:	$\int c_{3,,C1}$	=	$\{c_3\}$
$C_1(c_i) = 1$:	$\int c_{1,,C1} = \int c_{4,,C1}$	=	$\{c_1, c_4\}$
$C_1(c_i) = 2$:	$\int c_{2,,C1} = \int c_{5,,C1}$	=	$\{c_2, c_5\}$
$C_2(c_i) = 0$:	$\int c_{5,,C2}$	=	$\{c_5\}$
$C_2(c_i) = 1$:	$\int c_{2,,C2} = \int c_{3,,C2}$	=	$\{c_2, c_3\}$
$C_2(c_i) = 2$:	$\int c_{1,,C2} = \int c_{4,,C2}$	=	$\{c_1, c_4\}$
$C_3(c_i) = 0$:	$\int c_{1,,C3} = \int c_{4,,C3}$	=	$\{c_1, c_4\}$
$C_3(c_i) = 1$:	$\int c_{2,,C3}$	=	$\{c_2\}$
$C_3(c_i) = 2$:	$\int c_{3,,C3} = \int c_{5,,C3}$	=	$\{c_3, c_5\}$

The ordering that EVAL imposes on the candidate set will be defined in terms of these equivalence classes – i.e. EVAL does not order candidates directly, but rather orders equivalence classes of candidates. However, orderings are defined on elements of a set, and at this moment these equivalence classes do not form a set. The next step we need to accomplish is therefore to collect all of the equivalence classes into one set on which the ordering can then be defined. The set that has as its members all of the equivalence classes on some set A in terms of some equivalence relation R, is known as the quotient set of A modulo R (Enderton, 1977:58). We can define such a quotient set on the candidate set K modulo the equivalence relation $\approx_C$. The definition is given in

(15), and the quotient sets associated with each of the constraints in our example are given in (16).

(15) **<u>Def. 5</u>: Quotient set on *K* modulo $\approx_C$**

$K/C := \{[k]_{\approx_C} \mid k \in K\}$

(16) **Quotient sets on *K* modulo $\approx_C$ for each constraint**

$K/C_1 = \{\{c_3\}, \{c_1, c_4\}, \{c_2, c_5\}\}$

$K/C_2 = \{\{c_5\}, \{c_2, c_3\}, \{c_1, c_4\}\}$

$K/C_3 = \{\{c_2\}, \{c_1, c_4\}, \{c_3, c_5\}\}$

We are now finally in a position to define the ordering that EVAL imposes on (the quotient set on) the candidate set. In this ordering the equivalence class with the candidates that receive the smallest number of violations in terms of *C* will occupy the first position (will be the minimum, will precede all other equivalence classes), next will be the equivalence class with candidates that receive the second smallest number of violations in terms of *C*, etc.

(17) **<u>Def. 6</u>: The ordering relation $\leq_C$ on the set *K/C***

For all $C \in$ CON and all $[k_1]_{\approx_C}, [k_2]_{\approx_C} \in K/C$:

$[k_1]_{\approx_C} \leq_C [k_2]_{\approx_C}$ iff $C(k_1) \leq C(k_2)$.

Equivalence class $[k_1]_{\approx_C}$ "precedes" or "is better than" equivalence class $[k_2]_{\approx_C}$ if the candidates belonging to $[k_1]_{\approx_C}$ receives fewer violations in terms of *C* than the candidates belonging to $[k_2]_{\approx_C}$. The ordering $\leq_C$ that each of the constraints in our example imposes

on the candidate set is shown in (18). Note that this is exactly the same as the orderings shown in (8) above.

(18) **The $\leq_{Ci}$-ordering that EVAL imposes on each quotient set K/C_i**

C_1	C_2	C_3
$\{c_3\}$	$\{c_5\}$	$\{c_1, c_4\}$
|	|	|
$\{c_1, c_4\}$	$\{c_2, c_3\}$	$\{c_2\}$
|	|	|
$\{c_2, c_5\}$	$\{c_1, c_4\}$	$\{c_3, c_5\}$

The relation $\leq_C$ imposes an ordering on the quotient set *K/C* resulting in the ordered set $\langle K/C, \leq_C \rangle$ for each constraint. This set is the output of EVAL with respect to constraint *C*. Since the output of EVAL with respect constraints is defined in terms of the set *K/C* and ordering $\leq_C$, the characteristics of this set and ordering are important. In the next few sub-sections I discuss those characteristics of this set and ordering that are most directly relevant to our understanding of what a grammar is. First, I discuss the relation of individual candidates to the ordered set $\langle K/C, \leq_C \rangle$ (§2.2.1). Then I show that the ordering defined by $\leq_C$ is a chain (§2.2.2), and that this chain is guaranteed to always have a minimum (§2.2.3). Finally, I show that the set *K/C* is a partition on *K* (§2.2.4). The relevance of each of these results is discussed in the respective sections.

2.2.1 Individual candidates in the set $\langle K/C, \leq_C \rangle$

The ordering $\leq_C$ was defined in (17) above in terms of equivalence classes of candidates and not in terms of candidates. However, in actual practice we are usually interested in the relationship between candidates, and not between equivalence classes of candidates. In this section I will show that it is trivial matter to move from the ordering relationship

between equivalence classes of candidates to the relationship between individual candidates. It therefore does not matter whether we think of the ordering that EVAL imposes as an ordering on equivalence classes of candidates, or as an ordering on the individual candidates. I will first define an ordering on individual candidates in terms some constraint (this is the ordering that we are interested in when we actually do grammatical analyses), and then I will show that it is a trivial matter to move from the ordering on equivalence classes of candidates to this ordering on individual candidates.

(19) **<u>Def. 7</u>: The ordering relation $\leq_{C'}$ on the set *K***

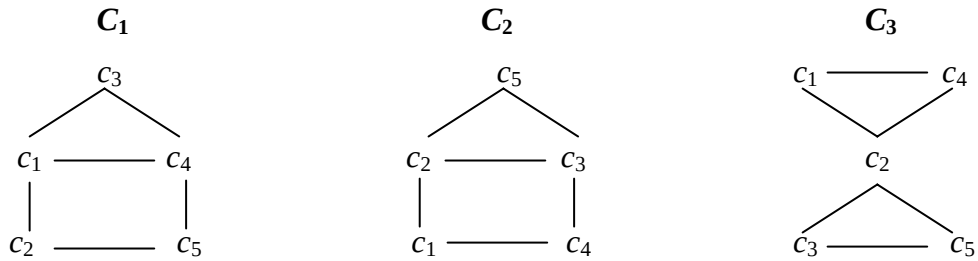
For all $C \in$ CON and all $k_1, k_2 \in K$:

$k_1 \leq_{C'} k_2$ iff $C(k_1) \leq C(k_2)$

The ordering $\leq_{C'}$ is intuitive – k_1 "precedes" or "is better than" k_2 if and only if k_1 receives fewer violations in terms C than k_2. The ordering $\leq_{C'}$ therefore defines a ordering on the candidate set K, resulting in the ordered set $\langle K, \leq_{C'}\rangle$. In (20) I give a graphic representation of the $\leq_{C'}$-orderings associated with each of the three constraints in our example. In this representation a candidate that appears higher precedes a candidate that appears lower – i.e. if k_1 appears higher than k_2, then $k_1 \leq_{C'} k_2$. If two candidates appear next to each other, then they are equal in terms of the ordering – i.e. if k_1 appears next to k_2, then $k_1 =_{C'} k_2$.

There is a natural relationship between the ordering $\leq_{C'}$ on candidates and the ordering $\leq_C$ on the equivalence classes – $k_1 \leq_{C'} k_2$ is only possible if $\mathfrak{f}k_{1,,C} \leq_C \mathfrak{f}k_{2,,C}$ and *vice versa*. That this is the case is obvious – these two orderings are defined by the same condition, i.e. if and only if $C(k_1) \leq C(k_2)$, then $k_1 \leq_{C'} k_2$ and $\mathfrak{f}k_{1,,C} \leq_C \mathfrak{f}k_{2,,C}$.

(20) **The $\leq_{c'}$-ordering that EVAL imposes on the candidate set for each constraint**



Because of this natural relationship between these two orderings it is a trivial matter to move from the ordered set $\langle K/C, \leq_C\rangle$ to the ordered set $\langle K, \leq_{C'}\rangle$. Suppose that we have only the ordered set $\langle K/C, \leq_C\rangle$, i.e. the set that contains not candidates but equivalence classes of candidates. Suppose further that we want to know how two candidates k_1 and k_2 are related to each other in terms of the ordering on individual candidates $\leq_{C'}$. All we need to do is to find the equivalence classes $\int k_{1,,C}$ and $\int k_{2,,C}$. If $\int k_{1,,C} <_C \int k_{2,,C}$, then also $k_1 <_{C'} k_2$. If $\int k_{1,,C} =_C \int k_{2,,C}$, then also $k_1 =_{C'} k_2$. If $\int k_{1,,C} >_C \int k_{2,,C}$, then also $k_1 >_{C'} k_2$. This can easily be proven formally. What we need is to show that there exists an order-embedding mapping from $\langle K/C, \leq_C\rangle$ to $\langle K, \leq_{C'}\rangle$. In (21) I first define what an order-embedding mapping is. In (22) I then define a mapping ψ from $\langle K/C, \leq_C\rangle$ to $\langle K, \leq_{C'}\rangle$. In (23) I show the result of applying ψ to the ordered sets $\langle K/C, \leq_C\rangle$ associated with each of the constraints in our example. Following that, I show that this mapping ψ is an order-embedding. The definition of an order-embedding is from Davey and Priestly (1990:10).

(21) **<u>Def. 8</u>: An order-embedding**

Let P and Q be ordered sets. A map $\varphi: P \rightarrow Q$ is said to be an order-embedding if $x \leq y$ in P iff $\varphi(x) \leq \varphi(y)$ in Q.

(22) **<u>Def. 9</u>: A mapping from $\langle K/C, \leq_C \rangle$ to $\langle K, \leq_{C'} \rangle$**

$\psi: \langle K/C, \leq_C \rangle \rightarrow \langle K, \leq_{C'} \rangle$ such that:

For all $\int k_{x,,C} \in K/C$ and for all $k_y \in \int k_{x,,C}$, $\psi(\int k_{x,,C}) = k_y$.[4]

(23) **Mapping equivalence classes onto their members**

a. **ψ applied to $\langle K/C_1, \leq_{C1} \rangle$:**

$\psi(\{c_3\}) = c_3$

$\psi(\{c_1, c_4\}) = c_1$ and c_4

$\psi(\{c_2, c_5\}) = c_2$ and c_5

b. **ψ applied to $\langle K/C_2, \leq_{C2} \rangle$:**

$\psi(\{c_5\}) = c_5$

$\psi(\{c_2, c_3\}) = c_2$ and c_3

$\psi(\{c_1, c_4\}) = c_1$ and c_4

c. **ψ applied to $\langle K/C_3, \leq_{C3} \rangle$:**

$\psi(\{c_2\}) = c_2$

$\psi(\{c_1, c_4\}) = c_1$ and c_4

$\psi(\{c_3, c_5\}) = c_3$ and c_5

We can show that the mapping ψ is an order-embedding mapping. This is stated as a theorem in (24).

---

[4] Note that ψ is not necessarily a function. ψ maps an equivalence class onto each of its members, and since an equivalence class can have more than one member, ψ can be a multi-valued mapping.

(24) **<u>Theorem 3</u>: That ψ is an order-embedding**

The mapping ψ as defined in Def. 9 (22) is an order-embedding.

<u>Proof of Theorem 3</u>: We need to show that $fk_{1„C} \leq_C fk_{2„C}$ if and only if $\psi(fk_{1„C}) \leq_{C'} \psi(fk_{2„C})$. First consider the *if* part – i.e. I will first show that $\psi(fk_{1„C}) \leq_{C'} \psi(fk_{2„C})$ implies $fk_{1„C} \leq_C fk_{2„C}$. Both $\psi(fk_{1„C})$ and $\psi(fk_{2„C})$ are members of the set $\langle K, \leq_{C'}\rangle$ – since ψ is defined as mapping into this set (Def. 9 (22)). If $\psi(fk_{1„C}) \leq_{C'} \psi(fk_{2„C})$, then by the definition of $\leq_{C'}$ (Def. 7 (19)) it follows that $C(\psi(fk_{1„C}) \leq C(\psi(fk_{2„C})$. But then by definition of the ordering $\leq_C$ (Def. 6 (17)), it follows directly that $fk_{1„C} \leq_C fk_{2„C}$.

Now consider the *only* part – i.e. I will now show that $fk_{1„C} \leq_C fk_{2„C}$ only if $\psi(fk_{1„C}) \leq_{C'} \psi(fk_{2„C})$. To show this, assume the opposite – i.e. assume that $fk_{1„C} \leq_C fk_{2„C}$ but that $\psi(fk_{1„C}) >_{C'} \psi(fk_{2„C})$.[5] Again, we know that both $\psi(fk_{1„C})$ and $\psi(fk_{2„C})$ are members of the set $\langle K, \leq_{C'}\rangle$ – since ψ is defined as mapping into this set (Def. 9 (22)). Based on the definition of $\leq_{C'}$ (Def. 7 (19)) we therefore know that if $\psi(fk_{1„C}) >_{C'} \psi(fk_{2„C})$, then $C(\psi(fk_{1„C}) > C(\psi(fk_{2„C})$. But based on the definition of $\leq_C$ (Def. 6 (17)), $C(\psi(fk_{1„C}) > C(\psi(fk_{2„C})$ implies $fk_{1„C} >_C fk_{2„C}$. And this contradicts the assumption that we started with, i.e. that $fk_{1„C} \leq_C fk_{2„C}$. So, this means that $fk_{1„C} \leq_C fk_{2„C}$ only if $\psi(fk_{1„C}) \leq_{C'} \psi(fk_{2„C})$.

□

This property of ψ is clear in the examples in (23). Consider C_1 as an example. We know from (18) that $\{c_3\} \leq_{C1} \{c_1, c_4\}$, and from (20) we know that $c_3 \leq_{C'1} c_1$, $c_3 \leq_{C'1} c_4$,

---

[5] If $\neg(\psi(fk_{1„C}) \leq_{C'} \psi(fk_{2„C})$ there are two possibilities to consider. It is possible that $\psi(fk_{1„C}) >_{C'} \psi(fk_{2„C})$ but it is also possible that $\psi(fk_{1„C})$ and $\psi(fk_{2„C})$ not comparable in terms of the relation $\leq_{C'}$. Because of the fact that constraints are functions, all candidates are comparable in terms of all constraints so that $\neg(\psi(fk_{1„C}) \leq_{C'} \psi(fk_{2„C})$ necessarily implies $\psi(fk_{1„C}) >_{C'} \psi(fk_{2„C})$.

and $c_1 \leq_{C'1} c_4$. In (23) we see that $\psi(\{c_3\}) = c_3$ and $\psi(\{c_1, c_4\}) = c_1$ and c_4. Therefore, we have $\{c_3\} \leq_{C1} \{c_1, c_4\}$ and $\psi(\{c_3\}) \leq_{C'1} \psi(\{c_1, c_4\})$. It can easily be checked that this true for all candidates and all three constraints.

What we have shown is that there exists an order-embedding mapping from $\langle K/C, \leq_C\rangle$ to $\langle K, \leq_{C'}\rangle$. It is therefore a straightforward matter to move from the ordered set of equivalence classes $\langle K/C, \leq_C\rangle$ to the ordered set of candidates $\langle K, \leq_{C'}\rangle$. Consequently, it does not matter in principle whether we think of the ordering that EVAL imposes on the candidate set in terms of individual candidates ($\leq_{C'}$) or in terms of equivalence classes of candidates ($\leq_C$) – the one can always be recovered from the other. In the rest of this chapter I will deal with the ordering only in terms of equivalence classes ($\leq_C$). The reason for this is that the ordering in terms of equivalence classes ($\leq_C$) is considerably simpler than the ordering in terms of individual candidates ($\leq_{C'}$). For one thing, the set K/C contains potentially fewer candidates than the set K – since every member of K/C can contain several members of K. There are therefore fewer ordering relations to consider in the set $\langle K/C, \leq_C\rangle$ than in the set $\langle K, \leq_{C'}\rangle$.[6]

This issue of the relationship between equivalence classes and individual candidates will also crop in a different guise towards the end of this chapter where I discuss the concept of "grammatical distinctness". Two candidates that belong to the

[6] There are more reasons. Strictly speaking $\leq_{C'}$ is not even an order, but only a "pre-order" or a "quasi-order" – i.e. it is a transitive and reflexive but not antisymmetric relation (Partee *et al*., 1993:207-208). Because of this many of the properties that we know to hold of orders in general do not necessarily hold of $\leq_{C'}$. Working with $\leq_C$, which is an order, is therefore more convenient – it allows us to use all of the standard concepts and theorems that apply to orders.

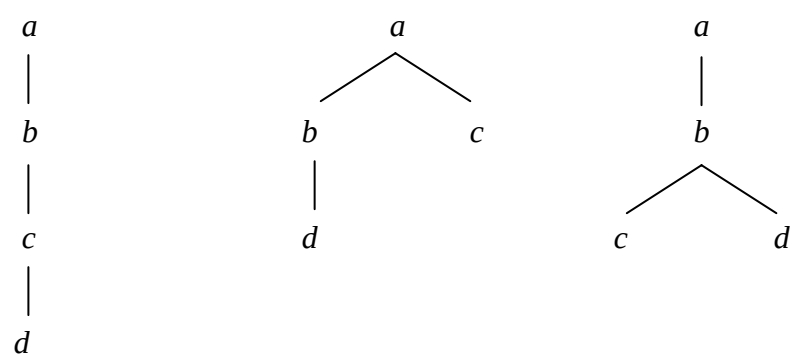
same equivalence class in terms of constraint *C* are "grammatically indistinct" in terms *C* (see §3.2.5 for more on this).

2.2.2 $\leq_C$ defines a chain

In this section I will show that the ordering that $\leq_C$ imposes on *K*/*C* is a chain ordering. Not all orderings are chain orderings. For an order to qualify as a chain ordering, it is necessary that any two elements be comparable. It is easiest to show the difference graphically. In (25) I show a graphic representation of three orderings on the set $A = \{a, b, c, d\}$ of which only the first is a chain.

(25) **Chains and non-chains**

A chain ordering **Non-chain I** **Non-chain II**



The second ordering in (25) is not a chain, because *b* and *c* are not comparable in this ordering, and neither is *c* and *d*. Similarly, the third ordering is not a chain since not all elements are comparable – here *c* and *d* are not comparable.

In (26) I formally define a chain (Davey and Priestley, 1990:3), and in (27) then state explicitly that $\leq_C$ defines a chain on *K*/*C*.

(26) **<u>Def. 10</u>: Definition of a chain**

Let *P* be an ordered set. Then *P* is a chain iff for all $x, y \in P$, either $x \leq y$ or $y \leq x$.

(27) **<u>Theorem 4</u>: That $\leq_C$ defines a chain**

The ordering that $\leq_C$ imposes on K/C is a chain.

<u>Proof of Theorem 4:</u> Consider $fk_{1„C}$, $fk_{2„C} \in K/C$, with $fk_{1„C}$ and $fk_{2„C}$ not necessarily distinct. By trichotomy of ≤ on ù (Enderton, 1977:62-63), it follows that one of the following is true: $C(k_1) < C(k_2)$, $C(k_1) > C(k_2)$, or $C(k_1) = C(k_2)$. By the definition of $\leq_C$ (Def. 6 (17)), it then follows directly that either $fk_{1„C} >_C fk_{2„C}$, $fk_{1„C} <_C fk_{2„C}$ or $fk_{1„C} =_C fk_{2„C}$. □

Looking back at the orderings in (18) that is associated with each of the three constraints in our example, it is clear that these three orderings are indeed chains. In each quotient set K/C_i all equivalence classes are comparable by the relation $\leq_{Ci.}$

What implications does this have for an OT grammar? It follows from this that there is no indeterminacy in the order that $\leq_C$ imposes on K/C. It is always possible to determine for any two equivalence classes in K/C how they are related with regard to each other in terms of $\leq_C$. And because of the natural relationship between $\langle K/C, \leq_C\rangle$ and $\langle K, \leq_{C'}\rangle$ (see §2.2.1 just above), it is also possible to determine for any two candidates how they are related to each other in terms of $\leq_{C'}$. For any constraint C and any two candidates k_1 and k_2, we can therefore determine from the set $\langle K/C, \leq_C\rangle$ whether $C(k_1) < C(k_2)$, $C(k_1) > C(k_2)$ or $C(k_1) = C(k_2)$.[7] This point will be discussed again when I consider the ordering imposed on the candidate set with respect to the full grammar (§3.2.2).

---

[7] In this sense the classic OT model, and specifically the model developed here, is different from OT with targeted constraints (Bakovic and Wilson, 2000, Wilson, 1999). A targeted constraint takes only a subset of the candidate set as domain. For a targeted constraint, EVAL can therefore impose an ordering only on those candidates in the domain of the constraint. Candidates not in the domain of a targeted constraint cannot be related to other candidates in terms of the ordering that EVAL imposes

2.2.3 The chain defined by $\leq_C$ always has a minimum

In this section I will show that the chain ordering that $\leq_C$ defines on the set K/C is guaranteed to have a minimum. The minimum will be that equivalence class that contains the candidates that receive the smallest number of violations in terms of C. The definition in (28) is from Davey and Priestley (1990:15).

(28) **<u>Def. 11</u>: Minimum of an ordered set**

Let P be an ordered set and $Q \subseteq P$. Then:

$a \in Q$ is the minimum of Q iff $a \leq x$ for every $x \in Q$.

(29) **<u>Theorem 5</u>: That $\langle K/C, \leq_C \rangle$ has a minimum**

The ordering $\leq_C$ always has a minimum in K/C.

<u>Proof of Theorem 5</u>:[8] This proof makes use of the notion of the *well-ordering* of ù under ≤ (Enderton, 1977:86). We say that ù is well-ordered under ≤, because every non-empty subset of ù is guaranteed to have a minimum under ≤.[9]

Constraints are functions with their ranges included in ù – that is for all candidates k and all constraints C, $C(k) \in$ ù. By the well-ordering of ù under ≤, it follows that there will be some candidate k such that C will map k onto a smaller number than all other candidates, i.e. there will be some k such that $C(k) \leq C(k')$ for all $k' \in K$. In

---

on the candidate set. The orderings associated with targeted constraints are therefore not chain orderings.

[8] A minimum is defined on a subset Q of the ordered set P. In the case under consideration here, the subset is equal to the superset – that is, K/C stands for both P and Q from the definition (which is possible since $K/C \subseteq K/C$). The proof therefore does not refer to the subset~superset relation.

[9] This is obviously true – in any set of natural numbers there will always be a smallest number. It can also be proved formally. See Enderton (1977:86-87) for a proof.

terms of the ordering $\leq_C$ (Def. 6 (17)) the equivalence class of k, $fk_{„C}$ (Def. 4 (13)), will then precede all other equivalence classes in the quotient set K/C (Def. 5 (15)), i.e. $fk_{„C} \leq_C fk'_{„C}$ for all $fk'_{„C} \in K/C$. And therefore $fk_{„C}$ is then the minimum in K/C. □

The orderings in (18) that are associated with each of the three constraints in our example clearly all have a minimum – in each of the orderings it is the equivalence class that appears highest on the graphic representations of the orderings. In this example this is every time the equivalence class containing the candidates that receive zero violations in terms of the specific constraint.

The converse of Theorem 5 is of course not necessarily true. Since there are constraints that can in principle assign an unbounded number of violations,[10] it is possible that the chain imposed by $\leq_C$ on K/C can be without a maximum. If this chain had no maximum and no minimum, then there would have been no uniquely identifiable point on the chain – for any member of the chain there would be infinitely many members above it and also infinitely members below it. The output of an OT grammar would not have been very informative had it been such an infinitely ascending and infinitely descending chain. We would never be able to refer uniquely to a specific level in the chain, and it would therefore not be possible to define access to the chain, i.e. access to the candidate set. I will return to this point again in §3.2.3 where I deal with the ordering imposed on the candidate set with reference to the full constraint set.

---

[10] There is nothing that limits the number of epenthetic segments in principle, so that DEP can assign an unbounded number of violations. There are also many markedness constraints that are not in principled limited in how many violations they can assign. Those markedness constraints that penalize a candidate for each instantiation of some marked structure (ONSET, NOCODA, *[+voice, +obstruent], etc.) can assign unboundedly many violations in principle – since unboundedly many violating structures can be inserted.

2.3.4 The set *K/C* is a partition on *K*

In this section I will show that the set K/C is a partition on K. What this means is that every member of K (every candidate) is included in one and exactly one of the equivalence classes contained in K/C. This is an important result for two reasons: First, since every candidate is included in some equivalence class, it means that the ordering $\leq_C$ does indeed give us information about every candidate. Secondly, every candidate is included in only one equivalence class, and every equivalence class can occupy only one position in the ordering that $\leq_C$ imposes on K/C. Because of the natural relationship between $\langle K/C, \leq_C \rangle$ and $\langle K, \leq_{C'} \rangle$ (see §2.2.1), this implies that every candidate is also guaranteed to occupy a unique slot in the ranking that EVAL imposes on the candidate set. The fact that K/C is a partition on K therefore assures that we can determine for every candidate what its unique relationship is to every other candidate in terms of the constraint C.

The definition of a partition in (30) is based on Enderton (1977:57) and Partee *et al*. (1993:46).

(30) **<u>Def. 12</u>: A partition**

A set P is said to be a partition on some set A iff:

(a) P consists of non-empty subsets of A.

(b) The sets in P are exhaustive – each element of A is in some set in P.

(c) The sets in P are disjoint – no two different sets in P have any element in common.

(31) **<u>Theorem 6</u>: *K/C* as a partition on *K***

K/C is a partition on *K*.

<u>Proof of Theorem 6</u>: Consider first (a), the requirement that the sets in *K/C* be non-empty. *K/C* is the quotient set on *K* modulo $\approx_C$ (Def. 5 (15)). Every member of *K/C* is therefore an equivalence class on *K* under the equivalence relation $\approx_C$ (Def. 4 (13)). Equivalence relations are reflexive (Def. 3 (11)), and therefore an equivalence class can never be empty. For any member $\mathit{f}k_{„C}$ of *K/C* it then follows that $\mathit{f}k_{„C}$ has at least one member, namely *k* (since $k \approx_C k$).

Now consider requirement (b), that the equivalence classes be exhaustive. Since constraints are functions with *K* as their domain ((3) and Theorem 1 (5)), it follows that $C(k)$ is defined for every $k \in K$. Then we have for every $k \in K$, $C(k) = C(k)$, and therefore $k \approx_C k$ (Def. 2 (9)). And finally we have for every $k \in K$, $k \in \mathit{f}k_{„C} \in K/C$ (Def. 4 (13) and Def. 5 (15)).

Now consider requirement (c), that the equivalence classes be disjoint. Let k_1, k_2, $k_3 \in K$ and let $\mathit{f}k_{2„C}$, $\mathit{f}k_{3„C}$ be equivalence classes associated with k_2 and k_3 respectively, as defined in Def. 4 (13). Now let $k_1 \in \mathit{f}k_{2„C}$ and $k_1 \in \mathit{f}k_{3„C}$. We need to show that $\mathit{f}k_{2„C} = \mathit{f}k_{3„C}$.

Since $k_1 \in \mathit{f}k_{2„C}$ and $k_1 \in \mathit{f}k_{3„C}$, we know that $k_2 \approx_C k_1$ and $k_3 \approx_C k_1$ (Def. 4 (13)). But $\approx_C$ is an equivalence relation (Theorem 2 (12)), and therefore $\approx_C$ is symmetric (Def. 3 (11)). Then we have $k_1 \approx_C k_3$, because of $k_3 \approx_C k_1$. But as an equivalence relation $\approx_C$ is also transitive (Def. 3 (11)). And therefore $k_2 \approx_C k_1$ and $k_1 \approx_C k_3$ implies $k_2 \approx_C k_3$. But again

since $\approx_C$ is transitive, it follows from $k_2 \approx_C k_3$ that for all k such that $k_3 \approx_C k$, also $k_2 \approx_C k$. And therefore for all k, if $k \in \mathfrak{f}k_{3,,C}$, then $k \in \mathfrak{f}k_{2,,C}$ (Def. 4 (13)). By similar reasoning we can show the converse, that for all k, if $k \in \mathfrak{f}k_{2,,C}$, then $k \in \mathfrak{f}k_{3,,C}$. Therefore we have $\mathfrak{f}k_{2,,C} = \mathfrak{f}k_{3,,C}$. □

In (16) the quotient sets associated with each of the constraints used in our examples were listed. Inspection of these equivalence classes will show that each of them are indeed partitions on K – every candidate is included in one and only one equivalence class in every quotient set.

We therefore now know that every candidate is included in exactly one of the equivalence classes that make up K/C. Together with the fact that $\leq_C$ defines a chain ordering on K/C and with the fact that there is a natural relationship between $\langle K/C, \leq_C \rangle$ and $\langle K, \leq_{C'} \rangle$ (see §2.2.1), this implies that every candidate has one unique spot in the ordering that EVAL imposes on the candidate set. From the information in the ordered set $\langle K/C, \leq_C \rangle$ we can determine for any two distinct candidates k_1 and k_2 how they are related to each other in terms of constraint C – i.e. whether they fare equally well on C ($C(k_1) = C(k_2)$), or whether one does better on C ($C(k_1) < C(k_2)$ or ($C(k_1) > C(k_2)$). There is no indeterminacy in the output of an OT grammar. I will return to this point again in §3.2.4 where the ordering imposed on the candidate set with reference to the full constraint set is discussed.

3. EVAL and the ordering associated with the full constraint set

Up to this point I have considered only how EVAL orders the candidate set with respect to individual constraints. Once EVAL has done this for every constraint in CON, we end up with as many orderings on the candidate set as there are constraints in CON. But the final output of the grammar is a single ordering on the candidate set. We therefore need a way in which these different orderings can be combined into one single ordering that corresponds to the whole grammar. This section of the chapter deals with how EVAL combines the orderings associated with different constraints.

To make the problem that needs to be addressed more concrete I will continue using the same example as earlier. I repeat the tableau from (6) in (32).

(32) **$\{c_1, c_2, c_3, c_4, c_5\}$ evaluated by $\|C_1 \circ C_2 \circ C_3\|$**

	C_1	C_2	C_3
c_1	*	**	
c_2	**	*	*
c_3		*	**
c_4	*	**	
c_5	**		**

In this grammar, c_3 is clearly the best candidate. If we remove c_3 from the candidate set, and consider only the four remaining candidates, then c_1 and c_4 tie as the best. If we also remove these candidates and consider only the two remaining candidates $\{c_2, c_5\}$, then c_5 is the best. In this grammar the candidates are therefore related as follows in terms of their harmony: $|c_3$ ™ $\{c_1, c_4\}$ ™ c_5 ™ $c_2|$. The ordering that EVAL imposes on the candidate set in terms of each individual constraint, and in terms of the grammar as a

whole can then be represented graphically as in (33). The orderings relative to the individual constraints are, of course, identical to the orderings shown above in (18).

(33) **Orderings on the candidate set relative to C_1, C_2, C_3, and $\|C_1 \circ C_2 \circ C_3\|$**

C_1	C_2	C_3	$\|C_1 \circ C_2 \circ C_3\|$
$\{c_3\}$	$\{c_5\}$	$\{c_1, c_4\}$	$\{c_3\}$
|	|	|	|
$\{c_1, c_4\}$	$\{c_2, c_3\}$	$\{c_2\}$	$\{c_1, c_4\}$
|	|	|	|
$\{c_2, c_5\}$	$\{c_1, c_4\}$	$\{c_3, c_5\}$	$\{c_5\}$
			|
			$\{c_2\}$

This makes it clear what we need: a way in which to combine the orderings associated with each of C_1, C_2 and C_3 such that the ordering that results is the ordering associated with $\|C_1 \circ C_2 \circ C_3\|$. Section §3.1 is dedicated to showing how this goal can be achieved. This combination process is defined in two stages. First, the Cartesian product is taken between the quotient sets associated with individual constraints (§3.1.1), and the set that results from this process is ordered lexicographically (§3.1.2). Then this new set and ordering is simplified by applying set intersection to each member of this new set (§3.1.3). The simplified set and ordering is the final output of the grammar, the final ordering that EVAL imposes on the candidate set for the grammar as a whole. Section §3.2 considers some of the properties of this set and ordering.

3.1 Combining the ordered sets associated with individual constraints

3.1.1 Taking the Cartesian product

We need a way in which to combine the ordered sets associated with each of the constraints. There are several different ways in which ordered sets can be joined – see Davey and Priestly (1990:17-19) for a discussion of the most important ways in which

this can be done. One way in which ordered sets can be combined is by taking the Cartesian product between them. When we take the Cartesian product between two sets a new set results that contains ordered pairs $\langle x_1, x_2 \rangle$ with x_1 coming from the first set and x_2 from the second set. The desirable property of this procedure is that a precedence relationship is established between the information coming from the two sets – the elements from the first set precede the elements from the second set in the ordered pairs $\langle x_1, x_2 \rangle$. In an OT grammar constraints are ranked and higher ranked constraints take precedence over lower ranked constraints. We have to take this fact in consideration when we combine the orderings associated with individual constraints into one conglomerate ordering for the grammar as a whole. The Cartesian product operation allows us to do exactly this – we take the Cartesian product of the ordered sets associated with each constraint in the order in which the constraints are ranked. This section explains how this can be achieved formally.

We usually think of the Cartesian product as the product between two sets. However, there is nothing that prohibits us from taking the Cartesian product of more than two sets. The Cartesian product of two sets is a set of ordered pairs $\langle x_1, x_2 \rangle$ with x_1 coming from the first set and x_2 from the second set. When we take the Cartesian between n sets, the result is a set of n-tuples $\langle x_1, x_2, \ldots x_n \rangle$ with x_i coming from the ith set. The definition of the Cartesian product in (34) is based on Enderton (1977:54). Simply restating this definition in terms of quotient sets associated with constraints, gives us the first step in the combination process. This is done in (35).

(34) **<u>Def. 13</u>: Cartesian product**

Let I be the set $\{1, 2, \ldots, n\}$, the *index set*, and let H be a function with domain I. Then, for each $i \in I$, we have a set $H(i)$. The Cartesian product of $H(i)$ for all $i \in I$ is defined as follows:

$$_{i \in I} H(i) := \{f \mid f \text{ is a function with domain } I \text{ and } \forall i\, (i \in I \rightarrow f(i) \in H(i))\}$$ [11]

(35) **<u>Def. 14</u>: Step 1 in combination process = Cartesian product between sets K/C_i**

Let I be the set $\{1, 2, \ldots, n\}$, the *index set*, such that $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$.

Let K/C_i be the quotient on K associated with C_i. We want the Cartesian product of all the quotient sets. We define this as follows:

$$_{i \in I} K/C_i := \{f \mid f \text{ is a function with domain } I \text{ and } \forall i\, (i \in I \rightarrow f(i) \in K/C_i)\}$$

The set $_{i \in I} K/C_i$ will be referred to as $K/C_\times$.

To see what the result of this process is, consider again the example from (6) and (32) above. In (16) the quotient sets associated with each of the three constraints were listed. In (36) I show the result of taking the Cartesian product between these three sets in the order $K/C_1 \times K/C_2 \times K/C_3$.

---

[11] Each $f \in {}_{i \in I} H(i)$ is of the following form: $f = \{\langle 1, x_1\rangle, \langle 2, x_2\rangle \ldots \langle n, x_n\rangle\}$ with $x_i \in H(i)$. The set that we actually want is a set whose members are ordered n-tuples, i.e. of the form $\langle x_1, x_2 \ldots x_n\rangle$ with $x_i \in H(i)$. However, it is a straightforward matter to uniquely match up every set $\{\langle 1, x_1\rangle, \langle 2, x_2\rangle \ldots \langle n, x_n\rangle\}$ with that n-tuple $\langle x_1, x_2 \ldots x_n\rangle$ to which it corresponds. The n-tuple that corresponds to the set $\{\langle 1, x_1\rangle, \langle 2, x_2\rangle \ldots \langle n, x_n\rangle\}$ is namely that n-tuple in which $x_i < x_j$ iff $\langle i, x_i\rangle, \langle j, x_j\rangle \in \{\langle 1, x_1\rangle, \langle 2, x_2\rangle \ldots \langle n, x_n\rangle\}$, and $i < j$. Since this uniquely matches up the sets $\{\langle 1, x_1\rangle, \langle 2, x_2\rangle \ldots \langle n, x_n\rangle\}$ with their corresponding n-tuples $\langle x_1, x_2 \ldots x_n\rangle$, we can without loss of exactness use the set notation and the n-tuple notation interchangeably.

(36) **Taking the Cartesian product of K/C_1, K/C_2 and K/C_3 from (17)**

$K/C_1 = \{\{c_3\}, \{c_1, c_4\}, \{c_2, c_5\}\}$

$K/C_2 = \{\{c_5\}, \{c_2, c_3\}, \{c_1, c_4\}\}$

$K/C_3 = \{\{c_1, c_4\}, \{c_2\}, \{c_3, c_5\}\}$

$K/C_\times =$

$$
\begin{array}{llll}
\{ & \langle\{c_3\}, \{c_5\}, \{c_1,c_4\}\rangle, & \langle\{c_3\}, \{c_5\}, \{c_2\}\rangle, & \langle\{c_3\}, \{c_5\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_3\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle, & \langle\{c_3\}, \{c_2,c_3\}, \{c_2\}\rangle, & \langle\{c_3\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_3\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle, & \langle\{c_3\}, \{c_1,c_4\}, \{c_2\}\rangle, & \langle\{c_3\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_1,c_4\}, \{c_5\}, \{c_1,c_4\}\rangle, & \langle\{c_1,c_4\}, \{c_5\}, \{c_2\}\rangle, & \langle\{c_1,c_4\}, \{c_5\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle, & \langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_2\}\rangle, & \langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle, & \langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_2\}\rangle, & \langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_2,c_5\}, \{c_5\}, \{c_1,c_4\}\rangle, & \langle\{c_2,c_5\}, \{c_5\}, \{c_2\}\rangle, & \langle\{c_2,c_5\}, \{c_5\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle, & \langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_2\}\rangle, & \langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle, \\
 & \langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle, & \langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_2\}\rangle, & \langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle \quad \}
\end{array}
$$

3.1.2 Imposing the lexicographic order on $K/C_\times$

At this point we have a set $K/C_\times$, but this is still an unordered set. In this section I will show how we can defined an ordering on this set. The lexicographic order is an ordering relationship on the product of *n* ordered sets that gives primacy to the order of the (*i*-1)th set over that of the *i*th set. This is desirable in OT – since the order imposed by the higher ranked constraints should be more important in the combined ordering.[12] The definition of the lexicographic ordering in (37) is based on Davey and Priestley (1990:19). However, they define the order only on a binary Cartesian product, while the definition in (37) is extended to cover arbitrary Cartesian products. In (38) I restate this definition in

[12] This follows from the strictness of strictness domination principle of OT (McCarthy, 2002b:4, Prince and Smolensky, 1993:78, 1997:1604). For more on this, see §3.2.1 below.

terms of the Cartesian product between the quotient sets associated with constraints – see (35) above.

(37) **<u>Def. 15</u>: Lexicographic order**

Let $_{i \in I} H(i)$ be the set as defined in Def. 13 (34) above, and let $\langle x_1, x_2, \ldots, x_n \rangle$, $\langle y_1, y_2, \ldots, y_n \rangle \in {}_{i \in I} H(i)$.

The lexicographic order on $_{i \in I} H(i)$ is defined as follows:

$\langle x_1, x_2, \ldots, x_n \rangle \leq \langle y_1, y_2, \ldots, y_n \rangle$ iff:

(i) For all $i \leq n$, $x_i = y_i$ (then $\langle x_1, x_2, \ldots, x_n \rangle = \langle y_1, y_2, \ldots, y_n \rangle$)

OR (ii) $\exists k$ such that:

- $\forall i\ (i < k \rightarrow x_i = y_i)$, and
- $x_k < y_k$ (then $\langle x_1, x_2, \ldots, x_n \rangle < \langle y_1, y_2, \ldots, y_n \rangle$)

(38) **<u>Def. 16</u>: Step 2 in the combination process = ordering $K/C_\times$**

Let $C_i \in$ CON, with the ranking $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$, and K/C_i the quotient set associated with constraint C_i (as defined in Def. 5 (15)). Let $\int x_{i,,Ci}, \int y_{i,,Ci} \in K/C_i$ be the equivalence classes of candidates x_i and y_i in terms of constraint C_i (as defined in Def. 4 (13)).

Let $\leq_{Ci}$ be the ordering that EVAL imposes on the candidate set in terms of constraint C_i (as defined in Def. 6 (17)).

Let $K/C_\times$ be the Cartesian product of K/C_i for all $i \in I$ (as defined in Def. 14 (35)).

Let $\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle, \langle \int y_{1,,C1}, \int y_{2,,C2} \ldots \int y_{n,,Cn} \rangle \in K/C_\times$.

Then $\leq_{\times,}$ the lexicographic order on $K/C_\times$, is defined as follows:

((38) continued)

$\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn} \rangle \leq_{\times} \langle f y_{1,,C1}, f y_{2,,C2} \ldots f y_{n,,Cn} \rangle$ iff:

(i) $\forall i (i \leq n \rightarrow f x_{i,,Ci} =_{Ci} f y_{i,,Ci})$

(then $\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn} \rangle =_{\times} \langle f y_{1,,C1}, f y_{2,,C2} \ldots f y_{n,,Cn} \rangle$)

OR (ii) $\exists k$ such that:

- $\forall i\ (i < k \rightarrow f x_{i,,Ci} =_{Ci} f y_{i,,Ci})$, and
- $f x_{k,,Ck} <_{Ck} f y_{k,,Ck}$.

(then $\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn} \rangle <_{\times} \langle f y_{1,,C1}, f y_{2,,C2} \ldots f y_{n,,Cn} \rangle$)

In order to make the discussion more concrete, I show in (39) the result of imposing the lexicographic order as defined here in (38) on the set $K/C_{\times}$ from (36).

As is clear from (36) and (39), the set $K/C_{\times}$ is a set of n-tuples of sets, while K/C_1, K/C_2 and K/C_3 are sets of sets. An OT grammar ultimately makes claims about (equivalence classes of) candidates,[13] and not about n-tuples of (equivalence classes of) candidates. It is therefore necessary to simplify the set $K/C_{\times}$ such that it is also a set of sets. The next sub-section introduces a method for doing this.

However, before we discuss this simplification process, I will first prove two lemmas about the ordering $\leq_{\times}$ on the set $K/C_{\times}$. The ordering that EVAL imposes on the simplified set (discussed in the next section) is defined in terms of the ordering $\leq_{\times}$. Although the characteristics of the ordering $\leq_{\times}$ are not themselves of direct relevance, these characteristics will become instrumental in the later discussion (§3.2.3 and §3.2.4).

[13] See §2.1.1 for the relationship between individual candidates and the equivalence classes of candidates.

(39) **Imposing the lexicographic order on $K/C_\times$**

$\langle K/C_1, \leq_{C1}\rangle$	$\langle K/C_2, \leq_{C2}\rangle$	$\langle K/C_3, \leq_{C3}\rangle$	$\langle K/C_\times, \leq_\times\rangle$
$\{c_3\}$	$\{c_5\}$	$\{c_1, c_4\}$	$\langle\{c_3\}, \{c_5\}, \{c_1,c_4\}\rangle$
$\{c_1, c_4\}$	$\{c_2, c_3\}$	$\{c_2\}$	$\langle\{c_3\}, \{c_5\}, \{c_2\}\rangle$
$\{c_2, c_5\}$	$\{c_1, c_4\}$	$\{c_3, c_5\}$	$\langle\{c_3\}, \{c_5\}, \{c_3,c_5\}\rangle$
			$\langle\{c_3\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle$
			$\langle\{c_3\}, \{c_2,c_3\}, \{c_2\}\rangle$
			$\langle\{c_3\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle$
			$\langle\{c_3\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle$
			$\langle\{c_3\}, \{c_1,c_4\}, \{c_2\}\rangle$
			$\langle\{c_3\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle$
			$\langle\{c_1,c_4\}, \{c_5\}, \{c_1,c_4\}\rangle$
			$\langle\{c_1,c_4\}, \{c_5\}, \{c_2\}\rangle$
			$\langle\{c_1,c_4\}, \{c_5\}, \{c_3,c_5\}\rangle$
			$\langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle$
			$\langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_2\}\rangle$
			$\langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle$
			$\langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle$
			$\langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_2\}\rangle$
			$\langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle$
			$\langle\{c_2,c_5\}, \{c_5\}, \{c_1,c_4\}\rangle$
			$\langle\{c_2,c_5\}, \{c_5\}, \{c_2\}\rangle$
			$\langle\{c_2,c_5\}, \{c_5\}, \{c_3,c_5\}\rangle$
			$\langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle$
			$\langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_2\}\rangle$
			$\langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle$
			$\langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle$
			$\langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_2\}\rangle$
			$\langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle$

(40) **<u>Lemma 1</u>: That $\leq_\times$ defines a chain**

$\leq_\times$ defines a chain on $K/C_\times$.[14]

[14] For a definition of a chain refer to Def. 10 (26) above.

Proof of Lemma 1: Let $\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn}\rangle, \langle f y_{1,,C1}, f y_{2,,C2} \ldots f y_{n,,Cn}\rangle \in K/C_{\times}$, K/C_i the quotient set associated with constraint C_i (Def. 5 (15)), and $\leq_{Ci}$ the ordering on this quotient set (Def. 6 (17)). Then we have $f x_{i,,Ci}, f y_{i,,Ci} \in K/C_i$ for all $i \leq n$ (Def. 14 (35)). But since $\leq_{Ci}$ defines a chain ordering on K/C_i (Theorem 4 (27)), we have that $f x_{i,,Ci} =_{Ci} f y_{i,,Ci}$ or $f x_{i,,Ci} >_{Ci} f y_{i,,Ci}$ or $f x_{i,,Ci} <_{Ci} f y_{i,,Ci}$ for all $i \leq n$. The ordering $\leq_{\times}$ is defined in terms of the orderings $\leq_{Ci}$, so that it follows immediately that any two elements in $K/C_{\times}$ are comparable, and therefore that $\leq_{\times}$ orders $K/C_{\times}$ in a chain. □

(41) **Lemma 2: That $\leq_{\times}$ always has a minimum**

The ordering $\leq_{\times}$ always has a minimum in $K/C_{\times}$.[15]

Proof of Lemma 2: There are two possible scenarios: (i) Either all candidates receive exactly the same number of violations in terms of all constraints, or (ii) there are at least two candidates that differ on at least one constraint. I will consider these two scenarios in turn below.

Scenario 1: All candidates receive exactly the same number of violations in terms of every constraint. If all candidates receive the same number of violations in terms some constraint C, then we have $k_i \approx_C k_j$ for all $k_i, k_j \in K$ (Def. 2 (9)). Then we have that all candidates will belong to the same equivalence class in terms of constraint C, i.e. for all k_i, $k_j \in K$ we have $k_i, k_j \in f k_{i,,C}$ (Def. 4 (13)). The quotient set on K associated with C, K/C (Def. 5 (15)), will then contain the set $f k_{i,,C}$. But since K/C is a partition on K (Theorem 6 (31)), its members exhaust all the candidates in k and are disjoint (Def. 12 (30)). Therefore, K/C contains only one member, namely $f k_{i,,C}$.

[15] For a definition of a minimum see Def. 11 (28) above.

Since all candidates receive the same number of violations in terms of all constraints, it follows that the quotient set associated with each constraint will have only one member, i.e. if there are n constraints, then for all $i \leq n$, K/C_i will contain only one member. If we take the Cartesian product of any number of sets each with only one member, the result will also be a set with only one member (Def. 13 (34)). The set $K/C_\times$ will therefore also contain only member. And then $\leq_\times$ defines a one-level chain, so that the single level on this chain is obviously the minimum of the chain.

Scenario 2: There are at least two candidates that differ on at least one constraint. I will assume that $\leq_\times$ does not have a minimum in $K/C_\times$, and then show that it leads to a contradiction.

Assume that $\leq_\times$ has no minimum in $K/C_\times$. Since $\leq_\times$ is a chain (Lemma 1 (40)), it then follows that $\leq_\times$ is an infinitely descending chain.

Let $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$ be the grammar under consideration. Then there must be a highest ranked constraint that does not rate all candidates equally (since under assumption there are at least two candidates that differ on at least one constraint). Call this constraint C_i. Then $\leq_{Ci}$ is the ordering associated with C_i as defined in Def. 6 (17), and K/C_i the quotient set associated with C_i as defined in Def. 5 (15). By assumption there are at least two candidates, k_1 and k_2, that are rated differently by C_i, i.e. $C_i(k_1) \neq C_i(k_2)$. From this it follows that $\leq_{Ci}$ establishes at least a two-level ordering on K/C_i. $\leq_{Ci}$ defines a chain (Theorem 4 (27)) with a minimum on K/C_i (Theorem 5 (29)). So, there must be some member $fx_{j,,Ci}$ of K/C_i such that $fx_{j,,Ci} <_{Ci} fk_{,,Ci}$ for all $fk_{,,Ci} \in K/C_i$ and $fk_{,,Ci} \neq fx_{j,,Ci}$ (i.e. $fx_{j,,Ci}$ is the minimum of $\geq_{Ci}$ in K/C_i).

All members of $K/C_\times$ are n-tuples of the form $\langle fx_{„C1}, fy_{„C2} \ldots fz_{„Cn}\rangle$ with $x, y, z \in K$ (Def. 14 (35)). Remember that C_i is the highest ranked constraint in terms of which any two candidates differ, and K/C_i is the quotient set associated with this constraint. Each n-tuple in $K/C_\times$ will have as its ith member a member from K/C_i. Also recall that $fx_{j„Ci}$ is the minimum of $\leq_{Ci}$ in K/C_i. Those n-tuples in $K/C_\times$ that have $fx_{j„Ci}$ as their ith member, will therefore be ordered before all n-tuples in $K/C_\times$ that have some other member of K/C_i as ith member in terms of the ordering $\leq_\times$ (Def. 16 (38)).

But by assumption $\leq_\times$ is an infinitely ascending chain. There must therefore be some n-tuple in $K/C_\times$ that is ordered before those n-tuples with $fx_{j„Ci}$ as their ith member. But this is not possible, since n-tuples with $fx_{j„Ci}$ as their ith member have just been shown to be ordered before all other n-tuples in $K/C_\times$.

Therefore, also in the second scenario $\leq_\times$ is guaranteed to have a minimum in $K/C_\times$. □

Now that these two lemmas have been proved, we can return to the main discussion. The next section shows how the set $K/C_\times$ and the ordering $\leq_\times$ can be simplified.

3.1.3 Simplifying the set $K/C_\times$ and the ordering $\leq_\times$

The ordering that EVAL will impose on the candidate set in our example is represented in (33). This ordering contains four members only, and each of its members is simply a set of candidates. This ordered set represents the final output of the grammar in our example. The ordered set $\langle K/C_\times, \leq_\times\rangle$ represented in (39) above is very different. It has 27

members, and its members are not simply sets of candidates, but ordered *n*-tuples of sets of candidates. We need to find a way to transform the set $\langle K/C_{\times}, \leq_{\times}\rangle$ into the set represented in (33). This section discusses the way in which we can simplify the set $K/C_{\times}$ and the ordering $\leq_{\times}$ into the set represented in (33). This transformation is defined in two steps.

I will first show how the set $K/C_{\times}$, a set of ordered *n*-tuples of sets, can be transformed into a set of sets. There is a straightforward way to create a set of sets of candidates by simplifying $K/C_{\times}$ – map each *n*-tuple from $K/C_{\times}$ onto the intersection between the members of the *n*-tuple. One limitation that needs to be placed on this mapping, is that it must not be able to map any element from $K/C_{\times}$ onto $\varnothing$. This mapping must therefore be undefined for elements of $K/C_{\times}$ where the intersection of the members of the *n*-tuple is $\varnothing$.[16] I will first define and illustrate this mapping. Later in §3.2 I will show that this mapping has the desired properties – i.e. that applying this mapping to $K/C_{\times}$ in (39) will indeed result in set represented in (33).

(42) **<u>Def. 17</u>: First half of step 3 in the combination process = *Intersect***

Let $K/C_{\times}$ be the set as defined in Def. 14 (35) above, and let $\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn}\rangle \in K/C_{\times}$. Then we define *Intersect*: $K/C_{\times} \rightarrow \wp(K)$ as follows:

Intersect$(\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn}\rangle)$ is undefined if $\int x_{1,,C1} \cap \int x_{2,,C2} \cap \ldots \cap \int x_{n,,Cn} = \varnothing$, and

Intersect$(\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn}\rangle) = \int x_{1,,C1} \cap \int x_{2,,C2} \cap \ldots \cap \int x_{n,,Cn}$ otherwise.

---

[16] The reason for this is that we want the new set of sets that results from this simplification to be a partition on the candidate set *K*, and the empty set cannot be a member of any partition – see Def. 12 (30). As for why the new set should be a partition on *K*, see §3.2.4 below.

Based on this mapping it is possible to define a new set that contains all of those sets onto which *Intersect* does indeed map some element of $K/C_{\times}$.

(43) **<u>Def. 18</u>: Collecting the output of *Intersect* into one set**

$K/C_{\text{Com}} := \{Z \mid\mid \exists\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn}\rangle \in K/C_{\times}$, such that

$$Z = \mathit{Intersect}(\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn}\rangle)\ \}$$

To see how this operation works, consider what it does to the set $K/C_{\times}$ from (36) and (39) above.

(44) ***Intersect* applied to $K/C_{\times}$**

$\mathit{Intersect}(\langle\{c_3\}, \{c_5\}, \{c_1,c_4\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_3\}, \{c_5\}, \{c_2\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_3\}, \{c_5\}, \{c_3,c_5\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_3\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_3\}, \{c_2,c_3\}, \{c_2\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_3\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle)$	$= \{c_3\}$
$\mathit{Intersect}(\langle\{c_3\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_3\}, \{c_1,c_4\}, \{c_2\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_3\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_5\}, \{c_1,c_4\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_5\}, \{c_2\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_5\}, \{c_3,c_5\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_2\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle)$	$= \{c_1, c_4\}$
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_2\}\rangle)$	is undefined
$\mathit{Intersect}(\langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle)$	is undefined

((44) continued)

$Intersect(\langle\{c_2,c_5\}, \{c_5\}, \{c_1,c_4\}\rangle)$	is undefined
$Intersect(\langle\{c_2,c_5\}, \{c_5\}, \{c_2\}\rangle)$	is undefined
$Intersect(\langle\{c_2,c_5\}, \{c_5\}, \{c_3,c_5\}\rangle)$	$= \{c_5\}$
$Intersect(\langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_1,c_4\}\rangle)$	is undefined
$Intersect(\langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_2\}\rangle)$	$= \{c_2\}$
$Intersect(\langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle)$	is undefined
$Intersect(\langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle)$	is undefined
$Intersect(\langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_2\}\rangle)$	is undefined
$Intersect(\langle\{c_2,c_5\}, \{c_1,c_4\}, \{c_3,c_5\}\rangle)$	is undefined

$\therefore\ K/C_{\text{Com}} = \{\{c_3\}, \{c_1, c_4\}, \{c_5\}, \{c_2\}\}$

The last thing that still needs to be done, is to define an ordering on this new set K/C_{Com}. This needs to be done with reference to the ordering $\leq_{\times}$ defined in Def. 16 (38) above.

(45) **<u>Def. 19</u>: Second half of Step 3 in the combination process: the ordering $\leq_{\text{Com}}$ on K/C_{Com}.**

Let $(fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}), (fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn}) \in K/C_{\text{Com}}$.

Then $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle, \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle \in K/C_{\times}$.

Then we define the order $\leq_{\text{Com}}$ on K/C_{Com} as follows:

$$(fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}) \leq_{\text{Com}} (fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn}) \text{ iff}$$

$$\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle \leq_{\times} \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle.$$

In (46) I show what the new ordered set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ looks like. Comparison with (33) will show that this is exactly the ordered set we were looking for.

(46) $\langle \boldsymbol{K/C_{\text{Com}}, \leq_{\text{Com}}} \rangle$

$$\begin{array}{c} \{c_3\} \\ | \\ \{c_1, c_4\} \\ | \\ \{c_5\} \\ | \\ \{c_2\} \end{array}$$

The set $\langle K/C_{\text{Com}}, \leq_{\text{Com}} \rangle$ is the final output of the grammar – this is rank-ordering that EVAL imposes on the candidate set. The properties of the set K/C_{Com} and the ordering $\leq_{\text{Com}}$ are therefore important and I will discuss them in §3.2 below. However, before we can do that, it is first necessary to prove a lemma about the mapping *Intersect*. The lemma will show that the inverse of *Intersect* is an order preserving mapping from $\langle K/C_{\text{Com}}, \leq_{\text{Com}} \rangle$ to $\langle K/C_{\times}, \leq_{\times} \rangle$. In and of itself this result is not of much interest. However, it will be used in §3.2 when the properties of $\langle K/C_{\text{Com}}, \leq_{\text{Com}} \rangle$ are discussed. Below I first give definitions for an inverse and an order preserving mapping, and then the lemma.

(47) **<u>Def. 20</u>: Inverse** (Enderton, 1977:44)

Let A and B be two sets, and R a relation on $A \times B$. R can then also be represented as set of ordered pairs, i.e. $R = \{\langle a_1, b_1 \rangle, \langle a_2, b_2 \rangle, \ldots\}$ with $a_i \in A$ and $b_i \in B$.

$R^{-1}: B \rightarrow A$, the inverse of R, is then a relation on $B \times A$, and is defined as follows:

$R^{-1} := \{\langle b, a \rangle \mid \langle a, b \rangle \in R\}$

(48) **<u>Def. 21</u>: The inverse of *Intersect***

Let $K' \subseteq K$. Then we can define $Intersect^{-1}$ as follows:

$Intersect^{-1}$ is a relation on $\wp(K) \times K/C_{\times}$ such that $Intersect^{-1}(K') = \langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle$ iff $Intersect(\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle) = K'$.

(49) **<u>Def. 22</u>: An order preserving mapping** (Davey and Priestley, 1990:10)

Let P and Q be ordered sets. A map φ: $P \rightarrow Q$ is said to order preserving if $x \leq y$ in P implies $\varphi(x) \leq \varphi(y)$ in Q.

(50) **<u>Lemma 3</u>: That *Intersect*$^{-1}$ is order preserving**

Intersect^{-1} is an order preserving mapping.

<u>Proof of Lemma 3</u>: Let $Intersect(\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn}\rangle) = K'$ and $Intersect(\langle f y_{1,,C1}, f y_{2,,C2} \ldots f y_{n,,Cn}\rangle) = K''$. Then $K', K'' \in K/C_{\text{Com}}$ (Def. 18 (43)). For *Intersect*$^{-1}$ to be order preserving, the following must therefore be true: If $K' \leq_{\text{Com}} K''$ then $Intersect^{-1}(K') \leq_{\times} Intersect^{-1}(K'')$. By substitution, this implies that for *Intersect*$^{-1}$ to be order preserving, $K' \leq_{\text{Com}} K''$ has to mean that also $\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn}\rangle \leq_{\times} \langle f y_{1,,C1}, f y_{2,,C2} \ldots f y_{n,,Cn}\rangle$. But this follows directly from the definition of $\leq_{\text{Com}}$ (Def. 19 (45)).

□

Consider again the example that we have been discussing all along. In (44) I showed the result of applying *Intersect* to the set $K/C_{\times}$. The set K/C_{Com} collects the output of *Intersect*. *Intersect*$^{-1}$ operates on K/C_{Com}. This is shown in (51).

(51) **Applying *Intersect*$^{-1}$ to K/C_{Com}**

$K/C_{\text{Com}} = \{\{c_3\}, \{c_1, c_4\}, \{c_5\}, \{c_2\}\}$

$Intersect^{-1}(\{c_3\}) \quad = \quad \langle\{c_3\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle$

$Intersect^{-1}(\{c_1, c_4\}) \quad = \quad \langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle$

$Intersect^{-1}(\{c_5\}) \quad = \quad \langle\{c_2,c_5\}, \{c_5\}, \{c_3,c_5\}\rangle$

$Intersect^{-1}(\{c_2\}) \quad = \quad \langle\{c_2,c_5\}, \{c_2,c_3\}, \{c_2\}\rangle$

From this example it is clear that $Intersect^{-1}$ in indeed order preserving. To see why, consider the first two mappings from (51). We know from (46) that $\{c_3\} \leq_{\text{Com}} \{c_1, c_4\}$. From (39) we know that $\langle\{c_3\}, \{c_2,c_3\}, \{c_3,c_5\}\rangle \leq_{\times} \langle\{c_1,c_4\}, \{c_1,c_4\}, \{c_1,c_4\}\rangle$, and therefore $Intersect^{-1}(\{c_3\}) \leq_{\times} Intersect^{-1}(\{c_1, c_4\})$. Therefore we have $\{c_3\} \leq_{\text{Com}} \{c_1, c_4\}$ and $Intersect^{-1}(\{c_3\}) \leq_{\times} Intersect^{-1}(\{c_1, c_4\})$. It can easily be checked that the same is true for all other comparisons between the mappings in (51). Now that we have shown that $Intersect^{-1}$ is an order preserving mapping, we can begin to consider the properties of the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$. This is done in the next section.[17]

3.2 The properties of K/C_{Com} and $\leq_{\text{Com}}$

The ordered set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ is the final output of the grammar – this is the ordering that EVAL imposes on the candidate set with respect to the full constraint ranking. The properties of this set are therefore important for our understanding of what an OT grammar is. This section is devoted to identifying those properties of this set that are most directly relevant to linguistic theory. First, it is necessary to confirm that this set does indeed agree with our intuitions about what the output of grammar looks like.[18] In particular, we need to confirm that the ordering on this set obeys the "strictness of strict

---

[17] The combination process of the orderings associated with individual constraints can be achieved in another way also. We can take the Cartesian product of the quotient sets associated with the two constraints that are ranked highest, i.e. $K/C_1 \times K/C_2$. The lexicographic order can then be imposed on this new set. The new ordered set $\langle K/C_1 \times K/C_2, \geq_{\times}\rangle$ can then be simplified according to the steps defined above – i.e. by applying *Intersect* to it, and by ordering the resulting set according to $\leq_{\text{Com}}$. This simplified ordered set then has exactly the same form as the quotient sets associated with individual constraints (it is a partition on K, ordered as a chain with a guaranteed minimum). This simplified set can then be combined with the quotient set associated with the third constraint in the hierarchy, K/C_3, in exactly the same manner as described for K/C_1 and K/C_2. This process is repeated recursively until the quotient sets associated with all of the constraints have been incorporated. The result of this process is provably equivalent to the process as I have defined it in the text.

[18] See the discussion on the introductory section to this chapter. We need to confirm that the explicatum agrees with our intuitions about the explicandum.

domination" principle. This is the focus of §3.2.1. After this some properties of the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ are discussed. Section §3.2.2 first shows that that $\leq_{\text{Com}}$ imposes a chain ordering on K/C_{Com}, and §3.2.3 then that this chain is guaranteed to have a minimum. Section §3.2.4 then shows that the set K/C_{Com} is a partition on the candidate set. Finally, §3.2.5 shows that the members of K/C_{Com} are equivalence classes on the candidate set K.

3.2.1 $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ and the strictness of strict domination

An OT grammar is a ranking between the constraints in CON. The candidate set is ordered relative to the constraint ranking with higher ranked constraints taking absolute precedence over lower ranked constraints. In particular, the highest ranked constraint that differentiates between two candidates takes precedence over all lower ranked constraints that also differentiate between these two candidates. Concretely, consider two candidates k_1 and k_2, and two constraints C_1 and C_2 ranked as $\|C_1 \circ C_2\|$. Suppose that C_1 prefers k_1 over k_2, but that C_2 prefers k_2 over k_1, i.e. $C_1(k_1) < C_1(k_2)$ but $C_2(k_1) > C_2(k_2)$. Because of the ranking $\|C_1 \circ C_2\|$, the ordering that C_1 imposes on k_1 and k_2 takes precedence – that is, for this mini-grammar as a whole, k_1 is more harmonic than k_2. This property of an OT grammar has been dubbed the "strictness of strict domination" or "strictness of domination" (McCarthy, 2002b:4, Prince and Smolensky, 1993:78, 1997:1604).

Since classic OT typically cares only about the relation between the winner and the mass of losers, strictness of domination is usually discussed as if it applies only to the relations between the winner and the losers. However, there is nothing that prohibits us from assuming that it also applies to the relationships between the losers, and that is in fact the assumption that I make in this dissertation. Strictness of domination therefore applies to the harmonic ordering relationships between any two candidates. In general

then, it is the highest ranked constraint in terms of which any two candidates are differentiated that determines the harmonic relation between them. For any two candidates k_1 and k_2, if C_1 is the highest ranked constraint that differentiates between them, then these two candidates will be harmonically ordered in terms of C_1. Stated differently, if $C_1(k_1) < C_1(k_2)$, then $|k_1$ ™ $k_2|$ even if there is some lower ranked constraint C_2 such that $C_2(k_1) > C2(k_2)$.

The set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ can only be viewed as the output of an OT grammar, if it can be shown that this set abides by the strictness of domination principle. This requirement is stated somewhat informally in (52).

(52) **Strictness of domination with reference to $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$, somewhat informally**

Let $K_1, K_2 \in K/C_{\text{Com}}$, with $k_1 \in K_1$ and $k_2 \in K_2$. Then:

$K_1 <_{\text{Com}} K_2$ iff the highest ranked constraint that distinguishes between k_1 and k_2 prefers k_1 over k_2.

The rest of this section is dedicated to showing that $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ does indeed abide by the strictness of domination principle. This is done in three steps: First, the concept "crucial constraint" is defined. The crucial constraint for two candidates is the highest ranked constraint that differentiates between them. Then, strictness of domination with reference to $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ is defined more precisely by using the concept of crucial constraints. Finally, it is proved that $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ abides by strictness of domination.

(53) **<u>Def. 23</u>: Crucial constraints**

Let k_1, $k_2 \in K$, and let the grammar under consideration be $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$.

Then we define $Crux_{1,2}$, the crucial constraint for k_1 and k_2, as follows:

$Crux_{1,2} = C_i$ such that $C_i(k_1) \neq C_i(k_2)$ and $\neg\exists j$ ($j < i$ and $C_j(k_1) \neq C_j(k_2)$).[19]

This definition identifies for any two candidates the highest ranked constraint that assigns a different number of violations to these two candidates. Now that the concept of a crucial constraint has been defined, we can state more formally what it would mean for the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ to abide by the strictness of domination principle. In (54) I first state the formal definition of strictness of domination. In (55) I then prove that $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ does indeed abide by this principle.

(54) **<u>Def. 24</u>: Strictness of domination with reference to $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$**

Let k_1, $k_2 \in K$, and K_1, $K_2 \in K/C_{\text{Com}}$ such that $k_1 \in K_1$ and $k_2 \in K_2$.

Let $fk_{i,,Cj}$ be the equivalence class of $k_i \in K$ in terms of constraint C_j as defined in Def. 4 (13) above, and $\leq_{Cj}$ the ordering associated with this constraint as defined in Def. 6 (17) above.

Let $Crux_{1,2}$ be the crucial constraint as defined just above in Def. 23 (53), and $\leq_{Crux1,2}$ the ordering that EVAL imposes on the candidate set with reference to $Crux_{1,2}$. Then:

$$K_1 <_{\text{Com}} K_2 \text{ iff } fk_{1,,Crux1,2} <_{Crux1,2} fk_{2,,Crux1,2}.$$

---

[19] Note that this definition implies that there will be no crucial constraints for two candidates that receive the same number of violations in terms of every constraint. This does not negate the result proved just below about the strictness of strict domination. Since all constraints agree on the relationship between two such candidates, it is not possible that a later constraint can disagree on an earlier constraint on the ordering relationship between two such candidates.

(55) **<u>Theorem 7</u>: Strictness of domination and $\langle K/C_{\mathrm{Com}}, \leq_{\mathrm{Com}}\rangle$**

$\langle K/C_{\mathrm{Com}}, \leq_{\mathrm{Com}}\rangle$ abides by strictness of domination.

<u>Proof of Theorem 7</u>: Let k_1, $k_2 \in K$, and K_1, $K_2 \in K/C_{\mathrm{Com}}$ such that $k_1 \in K_1$ and $k_2 \in K2$. Let $Crux_{1,2}$ be the crucial constraint for k_1 and k_2 as defined in Def. 23 (53), and $\leq_{Crux1,2}$ the ordering associated with this constraint as defined in Def. 6 (17).

We need to show that $K_1 <_{\mathrm{Com}} K_2$ iff $\mathit{f}k_{1,,Crux1,2} <_{Crux1,2} \mathit{f}k_{2,,Crux1,2}$. I will start by showing that $K_1 <_{\mathrm{Com}} K_2$ implies $\mathit{f}k_{1,,Crux1,2} <_{Crux1,2} \mathit{f}k_{2,,Crux1,2}$. The basic strategy in this proof is to work backwards through successive definitions until a statement can be made in terms of constraints. Since constraints have ù as range, it is easy to draw inferences on orderings associated with constraints.

Assume $K_1 <_{\mathrm{Com}} K_2$. By Def. 18 (43) and Def. 17 (42), the existence of K_1, $K_2 \in K/C_{\mathrm{Com}}$ implies the existence of some $\langle \mathit{f}x_{1,,C1}, \mathit{f}x_{2,,C2} \ldots \mathit{f}x_{n,,Cn}\rangle \in K/C_{\times}$ such that $K_1 = (\mathit{f}x_{1,,C1} \cap \mathit{f}x_{2,,C2} \cap \ldots \cap \mathit{f}x_{n,,Cn})$, and similarly the existence of some $\langle \mathit{f}y_{1,,C1}, \mathit{f}y_{2,,C2} \ldots \mathit{f}y_{n,,Cn}\rangle \in K/C_{\times}$ such that $K_2 = (\mathit{f}y_{1,,C1} \cap \mathit{f}y_{2,,C2} \cap \ldots \cap \mathit{f}y_{n,,Cn})$. From this is also follows that $\mathit{Intesect}^{-1}(K_1) = \langle \mathit{f}x_{1,,C1}, \mathit{f}x_{2,,C2} \ldots \mathit{f}x_{n,,Cn}\rangle$, and $\mathit{Intesect}^{-1}(K_2) = \langle \mathit{f}y_{1,,C1}, \mathit{f}y_{2,,C2} \ldots \mathit{f}y_{n,,Cn}\rangle$ (Def. 21 (48)). And since $\mathit{Intersect}^{-1}$ is order preserving (Lemma 3 (50)), $K_1 <_{\mathrm{Com}} K_2$ implies $\langle \mathit{f}x_{1,,C1}, \mathit{f}x_{2,,C2} \ldots \mathit{f}x_{n,,Cn}\rangle <_{\times} \langle \mathit{f}y_{1,,C1}, \mathit{f}y_{2,,C2} \ldots \mathit{f}y_{n,,Cn}\rangle$. By the definition of $\leq_{\times}$ (Def. 16 (38)), we then have that there is some i such that for all $j \leq i$, $\mathit{f}x_{j,,Cj} =_{Cj} \mathit{f}y_{j,,Cj}$ and $\mathit{f}x_{i,,Cj} <_{Ci} \mathit{f}y_{i,,Cj}$. Then by the definition of $\leq_{Ci}$ (Def. 6 (17)), it follows that there is some i such that for all $j \leq i$, $C_j(x_j) = C_j(y_j)$ and $C_i(x_i) < C_i(y_i)$. C_i is then the highest ranked constraint in terms of which the candidates in K_1 and K_2 differ.

Since $k_1 \in K_1$ and $k_2 \in K_2$ (by assumption), we therefore also have $C_i\ (k_1) < C_i\ (k_2)$. But since C_i is also the highest ranked constraint in terms of which k_1 and k_2 differ, it follows by Def. 23 (53) of crucial constraints that C_i is the crucial constraint for k_1 and k_2, i.e. $C_i = Crux_{1,2}$. Therefore we have $Crux_{1,2}(k_1) < Crux_{1,2}(k_2)$. But by the definition of the ordering $\leq_C$ (Def. 6 (17)) we then have $fk_{1,,Crux1,2} <_{Crux1,2} fk_{2,,Crux1,2}$. And therefore we have that $K_1 <_{\mathrm{Com}} K_2$ implies $fk_{1,,Crux1,2} <_{Crux1,2} fk_{2,,Crux1,2}$.

And the argument can be reversed to prove the converse. □

We can again check to confirm that this Theorem is indeed true of our example. In (56) I repeat from (39) the orderings $\leq_{Ci}$ associated with each of the three constraints, as well as the ordering $\leq_{\mathrm{Com}}$ from (46) associated with the grammar as a whole.

(56) **Orderings associated with constraints and the grammar as a whole**

$\langle K/C_1, \leq_{C1}\rangle$	$\langle K/C_2, \leq_{C2}\rangle$	$\langle K/C_3, \leq_{C3}\rangle$	$\langle K/C_{\mathrm{Com}}, \leq_{\mathrm{Com}}\rangle$
$\{c_3\}$	$\{c_5\}$	$\{c_1, c_4\}$	$\{c_3\}$
|	|	|	|
$\{c_1, c_4\}$	$\{c_2, c_3\}$	$\{c_2\}$	$\{c_1, c_4\}$
|	|	|	|
$\{c_2, c_5\}$	$\{c_1, c_4\}$	$\{c_3, c_5\}$	$\{c_5\}$
			|
			$\{c_2\}$

What we see is that the final ordering $\leq_{\mathrm{Com}}$ does not contradict the ordering $\leq_{C1}$ associated with the highest ranked constraint C_1. For instance, we have $\{c_3\} <_{C1} \{c_2, c_5\}$ which implies that $C_1(c_3) < C_1(c_5)$. On the other hand we have $\{c_5\} <_{C2} \{c_2, c_3\}$ which implies $C_2(c_5) < C_2(c_3)$. C_1 and C_2 therefore conflict in how they rate c_3 and c_5. However, since C_1 dominates C_2, the ordering $\leq_{\mathrm{Com}}$ associated with the grammar as a whole agrees with the ordering $\leq_{C1}$ associated with C_1 – i.e. $\{c_3\} <_{\mathrm{Com}} \{c_5\}$.

We have therefore now established that $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ abides by strictness of domination. This means that this set does indeed agree with our intuitions about what the output of an OT grammar should be like. This confirms that the procedures described above for arriving at this set, is an accurate depiction of what EVAL does to the candidate set. Now that we have established that the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ is of the correct form, we can investigate some its properties in more detail.

3.2.2 $\leq_{\text{Com}}$ defines a chain

In §2.2.2 I showed that the ordering that EVAL imposes on the candidate set with regard to each individual constraint is a chain ordering. In this section I will show that this also true of the ordering that EVAL imposes on the candidate set in terms of the grammar as a whole – i.e. not only is $\leq_C$ for every $C \in$ CON a chain, but so is the result of combining each of these orderings into the conglomerate ordering $\leq_{\text{Com}}$. For a discussion of what a chain see, refer to Def. 10 (26) above.

(57) **<u>Theorem 8</u>: That $\leq_{\text{Com}}$ defines a chain**

The ordering that $\leq_{\text{Com}}$ imposes on K/C_{Com} is a chain.

<u>Proof of Theorem 8</u>: The proof presented here follows the same basic strategy as the proof above for Theorem 7 – that is, substituting backwards through successive definitions. Let $K_1, K_2 \in K/C_{\text{Com}}$, with K_1 and K_2 not necessarily distinct.

By Def. 18 (43) and Def. 17 (42), the existence of $K_1, K_2 \in K/C_{\text{Com}}$ implies the existence of some $\langle f_{x_1,,C1}, f_{x_2,,C2} \ldots f_{x_n,,Cn}\rangle \in K/C_\times$ such that $K_1 = (f_{x_1,,C1} \cap f_{x_2,,C2} \cap \ldots \cap f_{x_n,,Cn})$, and similarly the existence of some $\langle f_{y_1,,C1}, f_{y_2,,C2} \ldots f_{y_n,,Cn}\rangle \in K/C_\times$ such that $K_2 = (f_{y_1,,C1} \cap f_{y_2,,C2} \cap \ldots \cap f_{y_n,,Cn})$.

Now, let $\leq_{Ci}$ be the ordering that EVAL imposes on the candidate set relative to constraint C_i (Def. 6 (17)). Since $\leq_{Ci}$ is a chain (Theorem 4 (27)), there are now three possible scenarios:

Scenario 1: For all $i \leq n$, $fx_{i,,Ci} =_{Ci} fy_{i,,Ci}$. Then, by the definition of $\leq_{\times}$ (Def. 16 (38)), it follows that $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle =_{\times} \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle$. This again implies by the definition of $\leq_{Com}$ (Def. 19 (45)), that $(fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}) =_{Com} (fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn})$, and therefore that $K_1 =_{Com} K_2$.

Scenario 2: There is some k such that $fx_{k,,Ck} >_{Ck} fy_{k,,Ck}$, and for all $i \leq k$, $fx_{i,,Ci} =_{Ci} fy_{i,,Ci}$. Then, by the definition of $\leq_{\times}$ (Def. 16 (38)), it follows that $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle >_{\times} \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle$, which again implies that $(fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}) >_{Com} (fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn})$ (Def. 19 (45)), and therefore that $K_1 >_{Com} K_2$.

Scenario 3: There is some k such that $fx_{k,,Ck} <_{Ck} fy_{k,,Ck}$, and for all $i \leq k$, $fx_{i,,Ci} =_{Ci} fy_{i,,Ci}$. Then, by the definition of $\leq_{\times}$ (Def. 16 (38)), it follows that $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle <_{\times} \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle$, which again implies that $(fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}) <_{Com} (fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn})$ (Def. 19 (45)), and therefore that $K_1 <_{Com} K_2$.

Finally, we then have that either $K_1 =_{Com} K_2$, or $K_1 >_{Com} K_2$, or $K_1 <_{Com} K_2$, and therefore that $\leq_{Com}$ defines a chain on K/C_{Com}. □

It is clear that this Theorem is true of the example that we have been using throughout the discussion. Referring back to the ordering $\langle K/C_{Com}, \leq_{Com}\rangle$ in (56) confirms that indeed any two elements on this ordering are comparable.

Recall that the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ is the final output of the grammar. This set represents the ordering that EVAL imposes on the candidate set with reference to the complete constraint hierarchy. The fact that this set is ordered as a chain is important. It shows that there is no indeterminacy in the output of the grammar. For any given candidate, it is possible to determine how it is harmonically related to any other candidate. Consider any two candidates k_1 and k_2. If k_1 and k_2 are both members of the same member of K/C_{Com}, i.e. $k_1, k_2 \in K_1 \in K/C_{\text{Com}}$, then k_1 and k_2 are equally harmonic. The grammar can then not distinguish between these two candidates, and they can be considered as grammatically indistinct.[20] However, if k_1 and k_2 belong to different members, K_1 and K_2, of K/C_{Com}, i.e. $k_1 \in K_1, k_2 \in K_2$ and $K_1 \neq K_2$, then either k_1 is more harmonic than k_2 or k_2 is more harmonic than k_1.

3.2.3 The chain defined by $\leq_{\text{Com}}$ always has a minimum

In §2.2.3 above I showed that the chain ordering $\leq_C$ associated with each individual constraint C is guaranteed to have a minimum. In this section I will show that the same is true for the ordering $\leq_{\text{Com}}$ (the ordering that EVAL imposes on the candidate set with reference to the whole grammar). For a definition of a minimum, see Def. 11 (28) above.

(58) **<u>Theorem 9</u>: That $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ always has a minimum**

The ordering $\leq_{\text{Com}}$ always has a minimum in K/C_{Com}.

<u>Proof of Theorem 9</u>: This theorem is proved by assuming the opposite, and then showing that this leads to a contradiction. In particular, I will show that assuming that

[20] See the discussion in §3.2.5 below about grammatical indistinctness.

$\leq_{Com}$ does not have a minimum in K/C_{Com} implies that the ordering $\leq_{\times}$ on $K/C_{\times}$ does not have a minimum, contra Lemma 2 (41).

Assume that $\leq_{Com}$ does not have a minimum in K/C_{Com}. We have established just above that $\leq_{Com}$ defines a chain on K/C_{Com} (Theorem 8 (57)). From this it follows that the only way in which $\leq_{Com}$ cannot have a minimum in K/C_{Com}, is if $\leq_{Com}$ defines an infinitely descending chain on K/C_{Com}. But if $\leq_{Com}$ defines an infinitely descending chain on K/C_{Com}, then for all $K_1 \in K/C_{Com}$, we have that there is some $K_2 \in K/C_{Com}$ such that $K_2 <_{Com} K_1$.

We have seen above that $Intersect^{-1}$ is an order preserving map between K/C_{Com} and $K/C_{\times}$ (Lemma 3 (50)). The ordering $\leq_{\times}$ on $K/C_{\times}$ will therefore inherit the properties of the ordering $\leq_{Com}$ on K/C_{Com}. It then follows that $\leq_{\times}$ on $K/C_{\times}$ is also an infinitely descending chain. But this contradicts Lemma 2 (41) which asserts that $\leq_{\times}$ is guaranteed to have a minimum in $K/C_{\times}$. □

The dual of this theorem is obviously not true – that is, it is not the case that the set $\langle K/C_{Com}, \leq_{Com}\rangle$ is guaranteed to have a maximum. There are constraints that can in principle assign an unbounded number of violations (such as DEP, ONSET, etc.).[21] Because of this, it is possible that $\leq_{Com}$ can define an infinitely ascending chain on K/C_{Com}.

[21] See footnote 10 above.

It is again clear that this Theorem is true of the example that we have been discussing throughout. $\{c_3\}$ in (56) precedes all other members of K/C_{Com} in terms of the ordering $\leq_{\text{Com}}$, and $\{c_3\}$ is therefore the minimum of the ordering $\leq_{\text{Com}}$ in the set K/C_{Com}.

Theorem 9 is a very important result. The set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ is the final output of the grammar. In order for this output to be in agreement with the way that we standardly think about an OT grammar, it should be possible to read off this set what the optimal candidate of classic OT is. This optimal candidate is the minimum of the set K/C_{Com} under the ordering $\leq_{\text{Com}}$. Since this minimum is guaranteed to exist it follows that we will always be able to determine what the optimal candidate of classic OT should be.

But there is another reason why it is important that the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ is guaranteed to have a minimum. From Theorem 8 (57) we know that $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ is a chain. If this chain had no minimum and no maximum, then it would have been impossible to identify individual members of this set with reference to the ordering on the set. For any member of this set there would then always have been another member that precedes it and another member that follows it. How would language users then access the information contained in this ordered set? There has to be a unique point on the ordering with regard to which the members of the set can be identified in terms of the ordering on the set. The minimum on the chain can serve this purpose. Since only one member of the set K/C_{Com} can be the minimum of this set under the ordering $\leq_{\text{Com}}$, we can uniquely identify this one member of K/C_{Com} in terms of the ordering $\leq_{\text{Com}}$. All other members of K/C_{Com} can then also be uniquely identified by stating their relationship to the minimum in terms $\leq_{\text{Com}}$. This minimum on the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$ serves as the point

through which language users can access the information contained in the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$.

3.2.4 The set K/C_{Com} is a partition on K

In §2.3.4 above I showed the quotient set K/C associated with every constraint C is a partition on the candidate set K. This means that every candidate $k \in K$ was contained in exactly one member of K/C. This result was important. The fact that every candidate is contained in a member of K/C implies that the ordered set $\langle K/C, \leq_C\rangle$ contains information about every candidate. The fact that every candidate is contained in only one member of K/C implies that every candidate can occur in only one place in the ordering $\leq_C$. The final output of the grammar is the set $\langle K/C_{\text{Com}}, \leq_{\text{Com}}\rangle$. In this section I will show that this set is also a partition on the candidate set K. This result will assure that also in this final output of the grammar every candidate is guaranteed to occupy exactly one position. For a definition of a partition, refer to Def. 12 (30) above.

(59) **<u>Theorem 10</u>: That K/C_{Com} is a partition on K**

K/C_{Com} is a partition on K.

<u>Proof of Theorem 10</u>: I will consider each of the three requirements for K/C_{Com} to be a partition on K in turn.

Consider first the requirement that K/C_{Com} consists of non-empty subsets of K. This follows directly from the definitions of K/C_{Com} (Def. 18 (43)) and *Intersect* (Def. 17 (42)). K/C_{Com} is defined such that all of its members are in the image of the set $K/C_{\times}$

under the relation *Intersect*.[22] And *Intersect* is defined as relation into $\wp(K)$ and such that $\varnothing$ is not in its range. We therefore have that all the members of K/C_{Com} are non empty subsets of K.

Now consider the second requirement, that the members K/C_{Com} be exhaustive subsets of K, i.e. that every $k \in K$ is a member of some member of K/C_{Com}. For all $C \in$ CON, let K/C be the quotient set associated with constraint C as defined in Def. 5 (15), and $\int x_{„C} \in K/C$, the equivalence class of x in terms of C (Def. 4 (13)). By Theorem 6 (31) we then have that K/C is a partition on K, and therefore for every $k \in K$ there is some $\int x_{„C} \in K/C$ such that $k \in \int x_{„C}$.

Now, let $K/C_{\times}$ be the Cartesian product over the quotient sets associated with each constraint (Def. 14 (35)), and $\langle \int x_{1„C1}, \int x_{2„C2} \ldots \int x_{n„Cn} \rangle \in K/C_{\times}$. Then for every $k \in K$ there is some $\langle \int x_{1„C1}, \int x_{2„C2} \ldots \int x_{n„Cn} \rangle \in K/C_{\times}$ such that $k \in \int x_{i„Ci}$ for all $i \leq n$. From the definition of *Intersect* (Def. 17 (42)), we then have that for every $k \in K$, there is some $\langle \int x_{1„C1}, \int x_{2„C2} \ldots \int x_{n„Cn} \rangle \in K/C_{\times}$ such that $k \in \mathit{Intersect}(\langle \int x_{1„C1}, \int x_{2„C2} \ldots \int x_{n„Cn} \rangle)$. And from the definition of K/C_{Com} (Def. 18 (43)) we have that $\mathit{Intersect}(\langle \int x_{1„C1}, \int x_{2„C2} \ldots \int x_{n„Cn} \rangle) \in K/C_{\text{Com}}$. Therefore, we have for every $k \in K$ that $k \in \mathit{Intersect}(\langle \int x_{1„C1}, \int x_{2„C2} \ldots \int x_{n„Cn} \rangle) \in K/C_{\text{Com}}$.

Now consider the third requirement, that the members of K/C_{Com} be disjoint subsets of K. I will show that any two members of K/C_{Com} that have an element in common are identical. Let $K_1, K_2 \in K/C_{\text{Com}}$ such $k \in K_1$ and $k \in K_2$. The existence of K_1,

---

[22] The image of set A under the relation F is defined as the set of all u such that there is some $v \in A$ such that $u = F(v)$ (Enderton, 1977:44), i.e. the image of A under $F := \{v \mid \exists u\ (u \in A \ \&\ F(u) = v)\}$.

K_2 implies the existence of $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle \in K/C_\times$ and $\langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle \in K/C_\times$ such that $Intersect(\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle) = K_1$ and $Intersect(\langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle) = K_2$ (Def. 17 (42) and Def. 18 (43)). Therefore $K_1 = fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}$, and $K_2 = fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn}$. But since $k \in K_1$ and $k \in K_2$ (by assumption), it follows that $k \in fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}$ and $k \in fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn}$. Therefore, for all $i \leq n$, $k \in fx_{i,,Ci}$ and $k \in fy_{i,,Ci}$.

But if $k \in fx_{i,,Ci}$, then $x_i \approx_{Ci} k$, and similarly if $k \in fy_{i,,Ci}$, then $y_i \approx_{Ci} k$ (Def. 4 (13)). But $\approx_{Ci}$ is an equivalence relation, and therefore symmetric and transitive (Theorem 2 (12)). Therefore, if $y_i \approx_{Ci} k$, then $k \approx_{Ci} y_i$, and if ($x_i \approx_{Ci} k$ and $k \approx_{Ci} y_i$), then $x_i \approx_{Ci} y_i$. From this it follows that $x_i \in fy_{i,,Ci,}$ so that ($x_i \in fy_{i,,Ci}$ and $x_i \in fx_{i,,Ci}$). But $fx_{i,,Ci}$ and $fy_{i,,Ci}$ are both elements of K/C_i, the quotient set associated with C_i (Def. 5 (15)). And we know that K/C_i is a partition on K (Theorem 6 (31)). The members of K/C_i are therefore disjoint, and therefore ($x_i \in fy_{i,,Ci}$ and $x_i \in fx_{i,,Ci}$) implies $fy_{i,,Ci} = fx_{i,,Ci}$. Then we have that $\langle fx_{1,,C1}, fx_{2,,C1} \ldots fx_{n,,Cn}\rangle = \langle fy_{1,,C1}, fy_{2,,C1} \ldots fy_{n,,Cn}\rangle$.

From $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle = \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle$ it follows that $Intersect(\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle) = Intersect(\langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle)$. And since $K_1 = Intersect\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle)$ and $K_2 = Intersect(\langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle)$, it follows that $K_1 = K_2$.

Finally, we then have that $k \in K_1$ and $k \in K_2$ implies $K_1 = K_2$, and therefore that the members of K/C_{Com} are disjoint. □

We can again check that this Theorem holds of the mini grammar that have been considering as an example throughout this chapter. The set K/C_{Com} for our mini grammar is $\{\{c_3\}, \{c_1, c_4\}, \{c_5\}, \{c_2\}\}$ (see (51)). In our example the candidate set K has only five members, i.e. $K = \{c_1, c_2, c_3, c_4, c_5\}$. It is clear that each member of K is included in exactly one member of $K/C_{\text{Com.}}$

The fact that K/C_{Com} is a partition on K is an important result. The ordered set $\langle K/C_{\text{Com,}} \leq_{\text{Com}}\rangle$ is the final output of the grammar. The fact that every candidate is included in one member of K/C_{Com} implies that every candidate is represented in the final output of the grammar. The fact that every candidate is included in only member of K/C_{Com} implies that every candidate occupies a unique place in the final output of the grammar. Together with the fact that $\leq_{\text{Com}}$ defines a chain on K/C_{Com} (Theorem 8 (57)), this means that there is no indeterminacy in the output of the grammar. It is always possible for any two candidates to determine how they are harmonically related to each other with regard to the ranking between the constraints. Consider $K_1, K_2 \in K/C_{\text{Com}}$, and $k_1 \in K_1$, $k_2 \in K_2$. If $K_1 =_{\text{Com}} K_2$, then we know that k_1 and k_2 are equally well-formed in terms of the grammar under consideration. If $K_1 <_{\text{Com}} K_2$, then we know that k_1 is more well-formed that k_2. If $K_1 >_{\text{Com}} K_2$, then we know that k_1 is less well-formed that k_2. Of particular relevance for the topic of this dissertation, is that this implies that any two candidates can be ordered with respect to each other – not only the best candidate in relation to the losers, but even any two losers in relation to each other.

3.2.5 The members of K/C_{Com} as equivalence classes on K

In §2 above I discussed the quotient sets K/C_i associated with the different constraints C_i. These quotient sets are not sets of candidates, but sets of sets of candidates. The set

K/C_{Com} is similar to these quotient sets. It is also a set of sets of candidates. However, there is an important difference between and K/C_{Com} and the sets associated with individual constraints. Let K/C_1 be the quotient set associated with constraint C_1, and $\{k_{1,,C1} \in K/C_1$. $\{k_{1,,C1}$ is therefore an equivalence class in terms of constraint C_1, implying that all candidates in $\{k_{1,,C1}$ receive the same number of violations in terms of C_1. If we have both k_1, $k_2 \in \{k_{1,,C1}$, then $C_1(k_1) = C_1(k_2)$. In terms C_1 these two candidates are indistinguishable. However it is possible that there is some other constraint in terms of which these two candidates differ – i.e. there could be some other constraint C_2 such that $C_2(k_1) \neq C_2(k_2)$. Although k_1 and k_2 are indistinguishable in terms of C_1 they might still be distinct from each other.

Now consider the set $K' \in K/C_{\text{Com}}$, and k_1, $k_2 \in K'$. The two candidates k_1 and k_2 are now completely indistinguishable in terms of the grammar – they receive exactly the same number violations in terms of every constraint. Two candidates that belong to same set of candidates in K/C_{Com} are therefore grammatically indistinct.[23] Although each member of K/C_{Com} can therefore contain more than one candidate, the grammar cannot distinguish between them. When we talk about a "candidate" in the output of an OT grammar, what we are actually referring to is rather a set of grammatically indistinct candidates. In this section I will prove that the sets in K/C_{Com} do indeed contain only grammatically indistinct candidates. In order to show this, I will first define an equivalence relation $\approx_{\text{Com}}$ on the candidate set, and then show that the members of K/C_{Com}

[23] Samek-Lodovici and Prince (1999) also use the concept of grammatical indistinctness. See also Hammond (1994, 2000) who uses this idea as a method of accounting for variation in the output – he assumes that variation arises when the set of best candidates has more than one member.

can be defined in terms of $\approx_{\text{Com}}$. For a definition of an equivalence relation, see Def. 3 (11) above.

(60) **<u>Def. 25</u>: $\approx_{\text{Com}}$ as a relation on *K***

Let the grammar under consideration be $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$, and $k_1, k_2 \in K$.

Then we define $\approx_{\text{Com}}$ as follows:

$$\approx_{\text{Com}} \subseteq K \times K \text{ such that } k_1 \approx_{\text{Com}} k_2 \text{ iff for all } i \leq n\ C_i(k_1) = C_i(k_2)$$

(61) **<u>Theorem 11</u>: That $\approx_{\text{Com}}$ is an equivalence relation**

$\approx_{\text{Com}}$ is an equivalence relation on K.

<u>Proof of Theorem 11</u>: $\approx_{\text{Com}}$ is by definition a binary relation on K. All that needs to be shown then is that it is reflexive, transitive and symmetric. Since it is defined in terms of constraints this is a rather straightforward matter. Constraints are functions into ù, and the relation = on ù is reflexive, transitive and symmetric. I consider each of the three requirements in turn below.

(i) That $\approx_{\text{Com}}$ is reflexive. Consider any candidate $k \in K$. Since $C(k) = C(k)$ for all $C \in$ CON, it follows from the definition of $\approx_{\text{Com}}$ (Def. 25 (60)) that $k \approx_{\text{Com}} k$. Therefore, $\approx_{\text{Com}}$ is reflexive.

(ii) That $\approx_{\text{Com}}$ is transitive. Consider any three candidates $k_1, k_2, k_3 \in K$ such that $k_1 \approx_{\text{Com}} k_2$ and $k_2 \approx_{\text{Com}} k_3$. By the definition of $\approx_{\text{Com}}$ (Def. 25 (60)) we then have that for all constraints $C \in$ CON, $C(k_1) = C(k_2)$ and $C(k_2) = C(k_3)$. But since = is transitive on ù, it follows from that $C(k_1) = C(k_3)$. From the definition of $\approx_{\text{Com}}$ it then follows that $k_1 \approx_{\text{Com}} k_3$. Therefore, $\approx_{\text{Com}}$ is transitive.

(iii) That $\approx_{Com}$ is symmetric. Consider any two candidates $k_1, k_2 \in K$ such that $k_1 \approx_{Com} k_2$. By the definition of $\approx_{Com}$ (Def. 25 (60)) we then have that for all constraints $C \in$ CON, $C(k_1) = C(k_2)$. But since = is symmetric on ù, it follows that $C(k_1) = C(k_2)$ implies $C(k_2) = C(k_1)$. And then again by the definition of $\approx_{Com}$ it follows that $k_2 \approx_{Com} k_1$. Therefore, $\approx_{Com}$ is symmetric. □

Now we can show that the members of the set K/C_{Com} can be fully defined in terms of the equivalence relation $\approx_{Com}$.

(62) **<u>Theorem 12</u>: The members of K/C_{Com} can be fully defined in terms of $\approx_{Com}$**

Let K be the candidate set, $K_1 \in K/C_{Com}$ and $k_1 \in K_1$. Then:

(i) $\forall k_2 \in K_1$, $k_1 \approx_{Com} k_2$, and

(ii) $\forall k_3 \in K$ such that $k_1 \approx_{Com} k_3$, $k_3 \in K_1$.

<u>Proof of Theorem 12</u>: Let $fk_{i,,Cj}$ be the equivalence class of candidate k_i in terms of constraint C_j (Def. 4 (13)). Now we can consider each of the two clauses of Theorem 12 in turn.

(i) We have by assumption that $k_1, k_2 \in K_1$. And the existence of K_1 implies that there is some n-tuple $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle \in K/C_{\times}$ such that $K_1 = Intersect(\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle) = fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}$ (Def. 17 (42) and Def. 18 (43)). And since $k_1, k_2 \in K_1$, it follows that $k_1, k_2 \in fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}$, and therefore that for all $i \leq n$, $k_1, k_2 \in fx_{,,Ci}$. By the definition of equivalence classes associated with individual constraints (Def. 4 (13)), it then follows that for all $i \leq n$, $C_i(k_1) = C_i(k_2)$. And then by the definition of the relation $\approx_{Com}$ (Def. 25 (60)) we have $k_1 \approx_{Com} k_2$.

(ii) We have by assumption that $k_1 \in K_1 \in K/C_{\text{Com}}$, $k_3 \in K$, and $k_1 \approx_{\text{Com}} k_3$. We need to show that then also $k_3 \in K_1$.

Consider first the assumption that $k_1 \in K_1 \in K/C_{\text{Com}}$. The existence of K_1 implies that there is some n-tuple $\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle \in K/C_{\times}$ such that $K_1 = \mathit{Intersect}(\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle) = \int x_{1,,C1} \cap \int x_{2,,C2} \cap \ldots \cap \int x_{n,,Cn}$ (Def. 17 (42) and Def. 18 (43)). And since by assumption $k_1 \in K_1$, it follows that for all $i \leq n$, $k_1 \in \int x_{i,,Ci}$, where $\int x_{i,,Ci}$ is the equivalence class of candidate x in terms of constraint C_i (Def. 4 (13)).

Now consider the assumption that $k_1 \approx_{\text{Com}} k_3$. From the definition of $\approx_{\text{Com}}$ (Def. 25 (60)), we have that for all $i \leq n$, $C_i(k_1) = C_i(k_3)$. From the definition of the relation $\approx_C$ (Def. 2 (9)) it then follows that for all $i \leq n$, $k_1 \approx_{Ci} k_3$. And since $\approx_{Ci}$ is an equivalence relation and therefore symmetric (Theorem 2 (12)), we also have $k_3 \approx_{Ci} k_1$. From the definition for equivalence classes associated with individual constraints (Def. 4 (13)), it then follows that for all $i \leq n$, $k_1, k_3 \in \int x_{i,,Ci}$.

The equivalence classes of each constraint are collected into a quotient set (Def. 5 (15)). Therefore for all $i \leq n$, we have that $\int x_{i,,Ci} \in K/C_{Ci}$. From Theorem 6 (31) we know that these quotient sets are partitions on K, and therefore that the members of the quotient sets are disjoint. The equivalence class that k_1 and k_3 belong to for each constraint, is therefore also the only equivalence class that each of them belongs to for that constraint. This means that the Cartesian product taken over the quotient sets of the different constraints, $K/C\times$, will contain one and only one n-tuple $\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle$ such that for all $i \leq n$, $k_1, k_3 \in \int x_{i,,Ci}$.

We have seen above that $k_1 \in K_1 \in K/C_{\text{Com}}$ implies the existence of some n-tuple $\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle \in K/C_{\times}$ such that for all $i \leq n$, $k_1 \in \int x_{i,,Ci}$. But since only one such an n-tuple exists for k_1, it is also the n-tuple for which it is true that for all $i \leq n$, $k_3 \in \int x_{i,,Ci}$. Then we have that for all $i \leq n$, $k_3 \in \int x_{1,,C1} \cap \int x_{2,,C2} \cap \ldots \cap \int x_{n,,Cn}$. And by the definition of *Intersect* (Def. 17 (42)) it follows that $k_3 \in \mathit{Intersect}(\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle)$. But as shown above, $K_1 = \mathit{Intersect}(\langle \int x_{1,,C1}, \int x_{2,,C2} \ldots \int x_{n,,Cn} \rangle)$. Therefore $k_3 \in K_1$.

Then we finally have that $k_1 \in K_1 \in K/C_{\text{Com}}$ and $k_1 \approx_{\text{Com}} k_3$, implies $k_3 \in K_1$. □

It is therefore possible to fully define the members of the set K/C_{Com} in terms of the equivalence relation $\approx_{\text{Com}}$.[24] This means that the candidates that are members of any given one of the sets in K/C_{Com} all have exactly the same number of violations in terms of every constraint. As far as the grammar is concerned, these candidates are all exactly the same. If our purpose is to describe the grammatical competence of the language user, this means that we can treat all candidates in each of the sets in K/C_{Com} alike, as if they are actually one single candidate – since according the grammar they are identical.

We can again check whether Theorem 12 is true of the example that we have been discussing throughout this chapter. The set K/C_{Com} in this example is $K/C_{\text{Com}} = \{\{c_3\}, \{c_1, c_4\}, \{c_5\}, \{c_2\}\}$ (see (51) above).There is only one member of this set that contains more than one candidate, namely $\{c_1, c_4\}$. If Theorem 12 is true of this example, then the two candidates c_1 and c_4 have to receive the same number of violations in terms of every constraint. Referring back to the tableau in (6) confirms that this is indeed true. We have

[24] The set K/C_{Com} is therefore also the quotient set of the relation $\approx_{\text{Com}}$ on the candidate set.

$C_1(c_1) = C_1(c_4) = 1$, $C_2(c_1) = C_2(c_4) = 2$, and $C_3(c_1) = C_3(c_4) = 0$. In this example the grammar can therefore not distinguish between these two candidates.

4. Summary

This chapter has investigated the formal properties of EVAL, the evaluative component of an OT grammar. The most important result of this chapter is that it established that a rank-ordering model of EVAL is entirely consistent with the basic architecture of a classic OT grammar. No changes or additions were made to the classic OT architecture in the development of this model. The rank-ordering model simply utilizes information that a classic OT grammar anyway generates. This chapter has developed a rigorous mathematical model of how the information generated by a classic OT grammar can be processed in order to establish a harmonic rank-ordering on the complete candidate set.

More specifically, this chapter has shown that: (i) The full candidate set is rank-ordered with respect to every constraint, and that this ordering is a chain with a minimum. (ii) The orderings of individual constraints can be combined to result in a final conglomerate ordering for the whole grammar. I showed that this conglomerate ordering agrees with our intuitions about the workings of an OT grammar (it abides by the principle of "strictness of strict domination"). The characteristics of this conglomerate ordering were investigated, and it was shown that it is an ordering on the full candidate set, that it is a chain ordering, and that it is guaranteed to have a minimum.

These results are important for our understanding of an OT grammar. Below I mention some of the most important perspectives on an OT grammar that this model

provides, and where appropriate I indicate how I use that result in the rest of the dissertation.

(i) *Richness of information*. The model developed in this chapter shows that a classic OT grammar generates a wealth of information about the relationships between the candidates. The information generated goes much further than simply distinguishing between the single optimal candidate and the losers. The grammar actually generates a much more detailed data structure that includes information about the harmonic relationship between any two candidates – i.e. also between candidates that are usually considered as losers.

An important question that this raises is whether this information is grammatically relevant. If the answer to this question is in the negative, then a classic OT grammar is much too powerful. It then massively over generates information. A more economic grammar would then have been one that would not have generated this information at all.

However, this chapter has shown that this information follows from the basic primitives of a classic OT grammar (constraint violation, constraint ranking, harmonic comparison between candidates). To formulate an alternative version of an OT grammar that does not generate this information will therefore be very difficult, if not impossible. This leads to the conclusion that it might be better to go the other direction – to assume that this information generated by the grammar is grammatically relevant.

This is the route that I take in this dissertation. I show several spheres in language performance where language users access and use this information. In variable phenomena language users access more candidates than just the single best candidate. They also access some of the candidates that are traditionally considered to be losers.

However, this implies that the “losers” cannot be a large amorphous group – then there would be no information on which language users can rely to select from among the “losers”. The information about the relationships between the losers is therefore crucial.

Also when rating the well-formedness of non-words language users access and use the richer information structure imposed by EVAL on the candidate set. Non-words that language users rate as more well-formed are simply non-words that correspond to candidates that occupy a higher slot in the rank-ordering that EVAL imposes on the candidate set. Similarly, I show that language users also use this information in lexical decision tasks. The lower a non-word occurs in the rank-ordering that EVAL imposes on the candidate set, the less seriously language users will consider the possibility that it is a word. Non-words corresponding to candidates lower in the rank-ordered candidate set, are therefore rejected quicker than non-words corresponding to candidates higher in the rank-ordered candidate set.

(ii) *No indeterminacy*. EVAL imposes two kinds of orderings on the candidate set – first EVAL orders the candidate set with regard to individual constraints (§2), and secondly EVAL orders the candidate set with regard to the whole grammar (§3). Both of these orderings have been shown to be on the whole candidate set (Theorem 6 (31) and Theorem 10 (59)). It has also been established that both of these orderings are chains (Theorem 4 (27) and Theorem 8 (57)). These two facts together imply that there is never indeterminacy in an OT grammar. It is always possible to determine for any two candidates how they are related to each other in terms of their harmony, whether with respect to an individual constraint or with respect to the whole grammar.

(iii) *A guaranteed output/unique access point*. The chain that EVAL imposes on the candidate set is guaranteed to have a minimum – again this holds of both the ordering with respect to individual constraints (Theorem 5 (29)) and with respect to the grammar as a whole (Theorem 9 (58)). Especially the fact that the chain associated with the whole grammar is guaranteed to have a minimum is relevant. This means that an OT grammar will always select a best candidate (or best set of candidates) from the candidate set. OT is therefore not a theory in which a derivation can crash because some input cannot be mapped onto any output candidate. OT is a forced-choice theory of grammar – the grammar is forced to select from the candidate set a best candidate.

There is another reason why it is important that the chain ordering imposed on the candidate set is guaranteed to have a minimum. This ensures that there is a uniquely identifiable point in terms of which access to the rank-ordered candidate set can be defined. A chain ordering with neither a maximum nor a minimum is both infinitely ascending and infinitely descending. If the chain ordering imposed by EVAL on the candidate set had neither a maximum nor a minimum, then no candidate could be uniquely identified with respect to its position in the chain ordering. For any candidate there would always be infinitely many candidates above it and infinitely many candidates below it. However, since the rank-ordered candidate set is guaranteed to have a minimum, there is a point on the candidate set that can be uniquely identified simply with respect to its position in the rank-ordering. In the rest of the dissertation I argue that the language user accesses the rank-ordered candidate set from its minimum. The minimum in the chain-ordering (the best candidate) is the most accessible, and the accessibility of candidates decreases the lower down they occur in the chain-ordering. In a variable

phenomenon the candidate that occupies the minimum position is therefore the most frequently observed output, the candidate that occupies the second slot is the second most frequently observed output, etc. In gradient well-formedness judgments, the candidate that occupies the minimum slot is rated best, and the further down from the minimum a candidate occurs, the less well-formed it is judged to be. In lexical decision tasks the candidate that occupies the minimum position in the chain-ordering is considered most seriously as a potential word, and is therefore associated with the slowest rejection times. The further down from the minimum a candidate occurs, the less seriously language users entertain the possibility that it might be a word, so that such candidates are associated with quicker rejection times.

(iv) *Independence from the candidate set*. The results of this chapter all assumed the existence of a set of candidates to be compared. However, the origin of this candidate set never featured in any of the proofs. The information structure that EVAL imposes on the candidate set therefore does not depend on the origin of the candidate set. EVAL can compare and order any set of candidate forms.

This result is also employed in the rest of the dissertation. We usually think of an OT grammar as comparing different output candidates for a single input – i.e. an input goes into GEN, and GEN then generates a set of candidate outputs for this input. However, the way in which EVAL works does not require the candidate set to originate in this manner. In particular, I argue that EVAL can compare candidates that are not morphologically related – that do not share the same input. GEN then generates a candidate set for several morphologically unrelated inputs, and we select subsets from these generated candidate sets to form a new set of candidates that is submitted to EVAL

for comparison (see Chapter 1 §1.2). A variable process does not always apply at the same rate in different contexts. In order to account for this, comparison across contexts is necessary, and I argue that this achieved by allowing EVAL to compare candidates from different inputs (from different contexts).

Also when non-words are compared, whether in a well-formedness judgment task or in a lexical decision task, the candidates that are compared are not candidates generated by GEN for some input. In accounting for these kinds of phenomena I also rely on the fact that EVAL can compare any set of candidate forms.

Appendix A: Definitions

(3) **Constraints as relations between the candidate set and ù**

Let CON be the universal set of constraints, and K the set of candidates to be evaluated. Then, $\forall C \in$ CON:

C: $K \rightarrow$ ù such that $\forall\ k \in K$, $C(k) =$ number of violations of k in terms of C

(4) **<u>Def. 1</u>: Functions**

A relation R from A to B is a function iff:

(a) the domain of $R = A$ (i.e. every member of A is mapped onto some member of B), and

(b) each element in A is mapped onto just one element in B (R is single valued).

(9) **<u>Def. 2</u>: The relation $\approx_C$ on K**

Let K be the candidate set to be evaluated by EVAL, and CON the set of constraints.

Then, for all $k_1, k_2 \in K$, and for all $C \in$ CON, let:

$$k_1 \approx_C k_2 \quad \text{iff} \quad C(k_1) = C(k_2).$$

(11) **<u>Def. 3</u>: An equivalence relation**

A binary relation R on some set is an equivalence relation on that set iff R is (i) reflexive, (ii) symmetric, and (iii) transitive.

(13) **<u>Def. 4</u>: Equivalence classes on K in terms of $\approx_C$**

For all $k_1 \in K$, $\{k_{1,,C} := \{k_2 \in K \mid k_1 \approx_C k_2\}$

(15) **<u>Def. 5</u>: Quotient set on K modulo $\approx_C$**

$K/C := \{\int k_{„C} \mid k \in K\}$

(17) **<u>Def. 6</u>: The ordering relation $\leq_C$ on the set K/C**

For all $C \in$ CON and all $\int k_{1„C}, \int k_{2„C} \in K/C$:

$\int k_{1„C} \leq_C \int k_{2„C}$ iff $C(k_1) \leq C(k_2)$.

(19) **<u>Def. 7</u>: The ordering relation $\leq_{C'}$ on the set K**

For all $C \in$ CON and all $k_1, k_2 \in K$:

$k_1 \leq_{C'} k_2$ iff $C(k_1) \leq C(k_2)$

(21) **<u>Def. 8</u>: An order-embedding**

Let P and Q be ordered sets. A map φ: $P \rightarrow Q$ is said to be an order-embedding if

$x \leq y$ in P iff $\varphi(x) \leq \varphi(y)$ in Q.

(22) **<u>Def. 9</u>: A mapping from $\langle K/C, \leq_C\rangle$ to $\langle K, \leq_{C'}\rangle$**

ψ:$\langle K/C, \leq_C\rangle \rightarrow \langle K, \leq_{C'}\rangle$ such that:

For all $\int k_{x„C} \in K/C$ and for all $k_y \in \int k_{x„C}$, $\psi\ (\int k_{x„C}) = k_y$.

(26) **<u>Def. 10</u>: Definition of a chain**

Let P be an ordered set. Then P is a chain iff for all $x, y \in P$, either $x \leq y$ or $y \leq x$.

(28) **<u>Def. 11</u>: Minimum of an ordered set**

Let P be an ordered set and $Q \subseteq P$. Then:

$a \in Q$ is the minimum of Q iff $a \leq x$ for every $x \in Q$.

(30) **<u>Def. 12</u>: A partition**

A set P is said to be a partition on some set A iff:

(a) P consists of non-empty subsets of A.

(b) The sets in P are exhaustive – each element of A is in some set in P.

(c) The sets in P are disjoint – no two different sets in P have any element in common.

(34) **<u>Def. 13</u>: Cartesian product**

Let I be the set $\{1, 2, \ldots, n\}$, the *index set*, and let H be a function with domain I. Then, for each $i \in I$, we have a set $H(i)$. The Cartesian product of $H(i)$ for all $i \in I$ is defined as follows:

$$_{i \in I} H(i) := \{f \mid f \text{ is a function with domain } I \text{ and } \forall i\, (i \in I \rightarrow f(i) \in H(i))\}$$

(35) **<u>Def. 14</u>: Step 1 in combination process = Cartesian product between sets K/C_i**

Let I be the set $\{1, 2, \ldots, n\}$, the *index set*, such that $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$.

Let K/C_i be the quotient on K associated with C_i. We want the Cartesian product of all the quotient sets. We define this as follows:

$$_{i \in I} K/C_i := \{f \mid f \text{ is a function with domain } I \text{ and } \forall i\, (i \in I \rightarrow f(i) \in K/C_i)\}$$

The set $_{i \in I} K/C_i$ will be referred to as $K/C_\times$.

(37) **<u>Def. 15</u>: Lexicographic order**

Let $_{i \in I} H(i)$ be the set as defined in Def. 13 (34) above, and let $\langle x_1, x_2, \ldots, x_n\rangle$, $\langle y_1, y_2, \ldots, y_n\rangle \in {}_{i \in I} H(i)$.

The lexicographic order on $_{i \in I} H(i)$ is defined as follows:

((37) continued)

$\langle x_1, x_2, \ldots, x_n\rangle \leq \langle y_1, y_2, \ldots, y_n\rangle$ iff:

(i) For all $i \leq n$, $x_i = y_i$ (then $\langle x_1, x_2, \ldots, x_n\rangle = \langle y_1, y_2, \ldots, y_n\rangle$)

OR (ii) $\exists k$ such that:

- $\forall i\ (i < k \rightarrow x_i = y_i)$, and
- $x_k < y_k$ (then $\langle x_1, x_2, \ldots, x_n\rangle < \langle y_1, y_2, \ldots, y_n\rangle$)

(38) **<u>Def. 16</u>: Step 2 in the combination process = ordering $K/C_\times$**

Let $C_i \in$ CON, with the ranking $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$, and K/C_i the quotient set associated with constraint C_i (as defined in Def. 5 (15)). Let $fx_{i,,Ci}, fy_{i,,Ci} \in K/C_i$ be the equivalence classes of candidates x_i and y_i in terms of constraint C_i (as defined in Def. 4 (13)).

Let $\leq_{Ci}$ be the ordering that EVAL imposes on the candidate set in terms of constraint C_i (as defined in Def. 6 (17)).

Let $K/C_\times$ be the Cartesian product of K/C_i for all $i \in I$ (as defined in Def. 14 (35)).

Let $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle, \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle \in K/C_\times$.

Then $\leq_\times$, the lexicographic order on $K/C_\times$, is defined as follows:

$\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle \leq_\times \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle$ iff:

(i) $\forall i(i \leq n \rightarrow fx_{i,,Ci} =_{Ci} fy_{i,,Ci})$

(then $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle =_\times \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle$)

OR (ii) $\exists k$ such that:

- $\forall i\ (i < k \rightarrow fx_{i,,Ci} =_{Ci} fy_{i,,Ci})$, and
- $fx_{k,,Ck} <_{Ck} fy_{k,,Ck}$.

(then $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle <_\times \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle$)

(42) **<u>Def. 17</u>: First half of step 3 in the combination process = *Intersect***

Let $K/C_\times$ be the set as defined in Def. 14 (35) above, and let $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle \in K/C_\times$. Then we define *Intersect*: $K/C_\times \rightarrow \wp(K)$ as follows:

Intersect($\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle$) is undefined if $fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn} = \varnothing$,

and *Intersect*($\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle$) = $fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}$ otherwise.

(43) **<u>Def. 18</u>: Collecting the output of *Intersect* into one set**

$K/C_{\text{Com}} := \{Z \mid\mid \exists\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle \in K/C_\times$, such that

$$Z = Intersect(\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle)\ \}$$

(45) **<u>Def. 19</u>: Second half of Step 3 in the combination process:**

the ordering $\leq_{\text{Com}}$ on K/C_{Com}.

Let $(fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn})$, $(fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn}) \in K/C_{\text{Com}}$.

Then $\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle$, $\langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle \in K/C_\times$.

Then we define the order $\leq_{\text{Com}}$ on K/C_{Com} as follows:

$$(fx_{1,,C1} \cap fx_{2,,C2} \cap \ldots \cap fx_{n,,Cn}) \leq_{\text{Com}} (fy_{1,,C1} \cap fy_{2,,C2} \cap \ldots \cap fy_{n,,Cn}) \text{ iff}$$

$$\langle fx_{1,,C1}, fx_{2,,C2} \ldots fx_{n,,Cn}\rangle \leq_\times \langle fy_{1,,C1}, fy_{2,,C2} \ldots fy_{n,,Cn}\rangle.$$

(47) **<u>Def. 20</u>: Inverse**

Let A and B be two sets, and F: $A \rightarrow B$ a relation on $A \times B$. F can then be represented as set of ordered pairs, $F = \{\langle a_1, b_1\rangle, \langle a_2, b_2\rangle, \ldots\}$ with $a_i \in A$ and $b_i \in B$. F^{-1}: $B \rightarrow A$, the inverse of F, is then a relation on $B \times A$, and is defined as follows:

$F^{-1} := \{\langle b, a\rangle \mid \langle a, b\rangle \in F\ \}$

(48) **<u>Def. 21</u>: The inverse of *Intersect***

Let $K' \subseteq K$. Then we can define $Intersect^{-1}$ as follows:

$Intersect^{-1}$: $\wp(K) \rightarrow K/C_{\times}$: $Intersect^{-1}(K') = \langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn} \rangle$ iff $Intersect(\langle f x_{1,,C1}, f x_{2,,C2} \ldots f x_{n,,Cn} \rangle) = K'$.

(49) **<u>Def. 22</u>: An order preserving mapping**

Let P and Q be ordered sets. A map φ: $P \rightarrow Q$ is said to order preserving if $x \leq y$ in P implies $\varphi(x) \leq \varphi(y)$ in Q.

(53) **<u>Def. 23</u>: Crucial constraints**

Let $k_1, k_2 \in K$, and let the grammar under consideration be $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$.

Then we define $Crux_{1,2}$, the crucial constraint for k_1 and k_2, as follows:

$Crux_{1,2} = C_i$ such that $C_i(k_1) \neq C_i(k_2)$ and $\neg\exists j$ ($j < i$ and $C_j(k_1) \neq C_j(k_2)$).

(54) **<u>Def. 24</u>: Strictness of domination with reference to $\langle K/C_{\mathrm{Com}}, \leq_{\mathrm{Com}} \rangle$**

Let $k_1, k_2 \in K$, and $K_1, K_2 \in K/C_{\mathrm{Com}}$ such that $k_1 \in K_1$ and $k_2 \in K_2$.

Let $f k_{i,,Cj}$ be the equivalence class of $k_i \in K$ in terms of constraint C_j as defined in Def. 4 (13) above, and $\leq_{Cj}$ the ordering associated with this constraint as defined in Def. 6 (17) above.

Let $Crux_{1,2}$ be the crucial constraint as defined just above in Def. 23 (53), and $\leq_{Crux1,2}$ the ordering that EVAL imposes on the candidate set with reference to $Crux_{1,2}$.

Then:

$$K_1 <_{\mathrm{Com}} K_2 \text{ iff } f k_{1,,Crux1,2} <_{Crux1,2} f k_{2,,Crux1,2}.$$

(60) **<u>Def. 25</u>: $\approx_{\mathbf{Com}}$ as a relation on *K***

Let the grammar under consideration be $\|C_1 \circ C_2 \circ \ldots \circ C_n\|$, and $k_1, k_2 \in K$.

Then we define $\approx_{\text{Com}}$ as follows:

$$\approx_{\text{Com}} \subseteq K \times K \text{ such that } k_1 \approx_{\text{Com}} k_2 \text{ iff for all } i \leq n\ C_i(k_1) = C_i(k_2)$$

Appendix B: Theorems and Lemmas

(5) **<u>Theorem 1</u>: Constraints as functions**

All constraints are functions.

(12) **<u>Theorem 2</u>: $\approx_C$ as an equivalence relation**

For all $C \in$ CON, $\approx_C$ is an equivalence relation on K.

(24) **<u>Theorem 3</u>: That ψ is an order-embedding**

The mapping ψ as defined in Def. 9 (16) is an order-embedding.

(27) **<u>Theorem 4</u>: That $\leq_C$ defines a chain**

The ordering that $\leq_C$ imposes on K/C is a chain.

(29) **<u>Theorem 5</u>: That $\langle K/C, \leq_C \rangle$ has a minimum**

The ordering $\leq_C$ always has a minimum in K/C.

(31) **<u>Theorem 6</u>: *K/C* as a partition on *K***

K/C is a partition on K.

(40) **<u>Lemma 1</u>: That $\leq_{\times}$ defines a chain**

$\leq_{\times}$ defines a chain on $K/C_{\times}$.

(41) **<u>Lemma 2</u>: That $\leq_{\times}$ always has a minimum**

The ordering $\leq_{\times}$ always has a minimum in $K/C_{\times}$.

(50) **<u>Lemma 3</u>: That *Intersect*$^{-1}$ is order preserving**

Intersect^{-1} is an order preserving mapping.

(55) **<u>Theorem 7</u>: Strictness of domination and $\langle K/C_{\mathrm{Com}}, \leq_{\mathrm{Com}}\rangle$**

$\langle K/C_{\mathrm{Com}}, \leq_{\mathrm{Com}}\rangle$ abides by strictness of domination.

(57) **<u>Theorem 8</u>: That $\leq_{\mathrm{Com}}$ defines a chain**

The ordering that $\leq_{\mathrm{Com}}$ imposes on K/C_{Com} is a chain.

(58) **<u>Theorem 9</u>: That $\langle K/C_{\mathrm{Com}}, \leq_{\mathrm{Com}}\rangle$ always has a minimum**

The ordering $\leq_{\mathrm{Com}}$ always has a minimum in K/C_{Com}.

(59) **<u>Theorem 10</u>: That K/C_{Com} is a partition on K**

K/C_{Com} is a partition on K.

(61) **<u>Theorem 11</u>: That $\approx_{\mathrm{Com}}$ is an equivalence relation**

$\approx_{\mathrm{Com}}$ is an equivalence relation on K.

(62) **<u>Theorem 12</u>: The members of K/C_{Com} can be fully defined in terms of $\approx_{\mathrm{Com}}$**

Let K be the candidate set, $K_1 \in K/C_{\mathrm{Com}}$ and $k_1 \in K_1$. Then:

(i) $\forall k_2 \in K_1$, $k_1 \approx_{\mathrm{Com}} k_2$, and

(ii) $\forall k_3 \in K$ such that $k_1 \approx_{\mathrm{Com}} k_3$, $k_3 \in K_1$.

CHAPTER 3

VOWEL DELETION IN LATVIAN

Vowels in final unstressed syllables are subject to variable deletion in Latvian (Karinš, 1995a).[1] The word *taka* "path" can therefore be pronounced as either [tá.ka] or [ták]. This deletion pattern is not random, but is strongly influenced by grammar. One of the most robust phonological factors that influence the likelihood of deletion is the length of the word – a vowel is more likely to delete from a tri-syllabic word than from a bi-syllabic word. In this chapter I provide an analysis of this variable deletion pattern within the rank-ordering model of EVAL. The main purpose of this chapter is not to give a full account of this phenomenon in Latvian, but rather to illustrate how the rank-ordering model of EVAL applies to variable processes. The focus of the discussion will therefore be more on the theoretical assumptions that I make than on the actual analysis of Latvian. Since the phenomenon of vowel deletion in Latvian is relatively simple, it lends itself to such a treatment. In the next two chapters I discuss two more complicated examples of variable phenomena (vowel deletion in Faialense Portuguese and final [t, d]-deletion in English). In these two later chapters, the focus is less on explaining how the rank-ordering model of EVAL works and more on the actual analysis of the phenomena.

The rest of this chapter is structured as follows: In §1 I introduce the data on variable vowel deletion in Latvian. An analysis of this phenomenon is then presented in §2. I postpone a discussion of alternative accounts of variation until the end of the section of the dissertation that deals with variable phenomena (Chapter 5 §3).

---

[1] I presented an earlier version of this chapter at the MIT Phonology Circle. The analysis presented here benefited greatly from discussions with the members of this group.

1. The data

Latvian builds syllabic trochees from the left edge of the word. The initial foot is the head foot so that stress falls on the initial syllable of a word (Karinš, 1995a, 1995b). The result of this is that in both bi-syllabic and tri-syllabic words, the final syllable is unstressed, and therefore occurs in a prosodically weak position.

(1) **Bi- and tri-syllabic words in Latvian**

(tá.ka)	'path'	(pá.zi)nu	'I knew'
(trá.ki)	'crazy'	(bá.lo)dis	'pigeon'

In the Riga dialect of Latvian all vowels in final unstressed syllables are subject to variable deletion. As a result of this, polysyllabic words in Latvian have two grammatical pronunciations. In (2) I give a few examples. These examples are from Karinš (1995a:18).

(2) **Variable deletion of final unstressed vowels in Latvian**

/spligti/	→	[spligti] ~ [spligt]	'dazzling' (m. pl.)
/pele/	→	[pele] ~ [pel]	'mouse'
/spligtas/	→	[spligtas] ~ [spligts]	'dazzling' (f. pl.)
/vajag/	→	[vajag] ~ [vaig]	'need'
/sakne/	→	[sák.ne]~[sa.kn̥]	'root' (f. sg.)

This deletion process is not random, but is strongly influenced by phonology. Karinš (1995a) identifies several phonological factors that co-determine the likelihood that a vowel will delete. One of the most significant is the length of the word – a vowel is more likely to delete from a tri-syllabic than from a bi-syllabic word. I will focus on this

aspect of the phenomenon here. The table in (3) contains the data on the likelihood of deletion from bi- and tri-syllabic words. The data are from Karinš (1995a:21).

(3) **Variable vowel deletion in bi- and tri-syllabic words in Latvian**[2, 3]

	n	**Retained**	**Deleted**	**% deleted**
Bi-syllabic	1,389	264	1,125	81
Tri-syllabic	743	59	684	92

In this dissertation I propose two changes in the way that we think about EVAL. First, I propose that EVAL imposes a rank-ordering on the full candidate set and does not distinguish only between the winner and the mass of losers. Secondly, I propose that EVAL can compare morphologically unrelated forms with each other. In the rest of this chapter I will show how we can use these two extensions to EVAL to account for the variable deletion pattern exemplified in (3).

2. The analysis

There are three aspects of the variation pattern in (3) that we need to account for. These are: (i) *Intra-contextual variation*. For both a bi-syllabic and a tri-syllabic input there are two variants – one with deletion of the final unstressed vowel and one with retention of this vowel. Of these two variants the deletion variant is the more frequently observed. This is what I will refer as "intra-contextual variation". We need to account for the fact that for any given input one of the variants is observed more frequently than the other. I

[2] These data are based on recordings of 8 Latvian speakers that Karinš made in Riga in 1991.

[3] Karinš (1995a:27) reports that the difference between the deletion rates in these two classes was significant for all of his subjects.

discuss this aspect of the variation pattern in §2.1. (ii) *Inter-contextual variation*. Although deletion is the most frequent variant for both bi-syllabic and tri-syllabic forms, deletion is observed more frequently in tri-syllabic contexts than in bi-syllabic contexts. This variation between contexts is discussed in §2.2. (iii) *Limits on variation*. Only two variants are observed – either retention or deletion of the vowel. But there are more ways than just deletion to avoid a final unstressed vowel – for instance, it is also possible to assign secondary stress to the final vowel. We need to explain why other variants such as these are never observed. This is discussed in §2.3. This chapter then ends with a discussion of ranking arguments (McCarthy, 2002b:4-5) in §2.4. Under the rank-ordering model of EVAL, the notion of a ranking argument has to be redefined. Finally, I summarize the analysis in §2.5.

2.1 Intra-contextual variation: more deletion than retention

My basic argument is the following: EVAL imposes a harmonic ordering on the full candidate set, and language users have access to all levels of this fully ordered set. However, the accessibility of a candidate is directly related to how high it occurs on this harmonic ordering. The candidate rated best by EVAL (occurring highest in the harmonic ordering = the optimal candidate of classic OT), is the most accessible; the candidate rated second best is the second most accessible; etc.

Returning to the Latvian example: Vowels from final unstressed syllables can be realized in one of two ways. They can be faithfully preserved in pronunciation, or they can delete. As the data in the table in (3) show, the deletion candidate is observed more frequently than the faithful candidate in both bi- and tri-syllabic words. For both of these two kinds of words EVAL therefore has to rate the deletion candidate better than the

retention candidate, i.e. |∅ ™ v̌|.[4] This can be achieved if the highest ranked constraint that disfavors the retention candidate outranks the highest ranked constraint that disfavors the deletion candidate.[5] At this point it is therefore necessary to determine the constraints that are violated by the retention and the deletion candidates. For the deletion candidate, this is easy – it violates the anti-deletion constraint MAX. I propose that the constraints that drive deletion are different for the bi-syllabic and the tri-syllabic forms. Deletion in bi-syllabic forms is driven by a constraint against vowels in unstressed final syllables. Deletion in tri-syllabic forms is driven by a constraint that requires all syllables to be parsed into higher prosodic structure. I discuss the bi-syllabic forms first, and I begin by defining the constraint that drives deletion in these forms in (4).

(4) ***v̌]$_{\sigma}$]$_{\omega}$**

Do not allow a vowel in an unstressed prosodic word-final syllable.

It is well established that vowels are marked/avoided in prosodically weak positions. This is often the driving force behind vowel reduction and deletion – see for instance Crosswhite (2000a, 2000b, 2001) and the discussion of Faialense Portuguese in Chapter 4 of this dissertation. In a language like Latvian with initial stress, the word final syllable is arguably the prosodically weakest position. It is therefore not surprising that this is the position in which vowels are deleted.

---

4 Throughout the discussion I will use ∅ to stand for the deletion candidate and [v̌] for the retention candidate.

5 The idea of constraints favoring or disfavoring a candidate comes from Samek-Lodovici and Prince (1999). Let $C(x)$ represent the number of violations constraint C assigns to candidate x, and K the set of all candidates under consideration. For some candidate *Can* to be favored by constraint C, the following statement must then be true: $\neg\exists\ k \in K\ (C(k) < C(Can))$. Conversely, for some candidate *Can* to be disfavored by constraint C, the following statement must be true: $\exists\ k \in K\ (C(k) < C(Can))$.

We have to answer two questions about the constraint in (4). (i) Why is this constraint formulated as a constraint against vowels in unstressed, prosodic word-final syllables? Why is it not rather formulated as a constraint against unstressed vowels in prosodic word-final position? There are several examples of languages that delete vowels from absolute final position in prosodic words – see the discussion in Chapter 4 §3.2.2. The reason for this is that Latvian deletes vowels even from closed final unstressed syllables. See for instance the word *spligtas* 'dazzling' (f. pl.) from (2) above. The vowel from the final syllable of this word is also subject to deletion, so that this word can be pronounced as [splígts]. It is therefore necessary that the operative constraint is violated also by an unstressed vowel that is not in absolute final position in a prosodic word. (ii) Why do we not simply state this as a constraint against final unstressed syllables? The reason for this is that the deletion of the vowel from the final unstressed syllable does not always result in deletion of the final syllable. When the unstressed vowel is followed by a nasal or liquid, the nasal or liquid becomes syllabic so that syllable count of the word does not change. See for instance the word *sakne* 'root' from (2) above. The final nasal of this word syllabifies when the vowel deletes so that it is pronounced as [sa.kn̩].[6]

The retention candidate therefore violates the markedness constraint $*\check{v}]_{\sigma}]_{\omega}$ and the deletion candidate violates MAX. Under the ranking $||*\check{v}]_{\sigma}]_{\omega} \circ \text{MAX}||$,[7] the deletion candidate violates the lower ranking constraint and will therefore be rated as better by

[6] This is echoed in English – vowels from unstressed word final syllables often delete when followed by liquids/nasals. The liquid/nasal then becomes syllabic. See for instance *battle* pronounced as [bætl̩] and *even* pronounced as [i:vn̩].

[7] On the motivation for this ranking, see the discussion on ranking arguments in §2.4 below.

EVAL. This is shown in the tableaux in (5). On the typographical conventions used in the tableau, see Chapter 1 §2.2.1.

(5) **Deletion preferred over retention in bi-syllabic forms**

			* v̌]$_{\sigma}$]$_{\omega}$	MAX
/taka/	2	(tá.ka)	*	
	1	(ták)		*

Output of EVAL

ták $_{\text{MAX}}$

|

tá.ka $_{*\check{v}]\sigma]\omega}$

The deletion candidate violates MAX, and the retention candidate violates *v̌]$_{\sigma}$]$_{\omega}$. However, because of the ranking ||*v̌]$_{\sigma}$]$_{\omega}$ o MAX||, EVAL rates the deletion candidate better than the retention candidate. EVAL therefore imposes the following rank-ordering on the candidate set: |∅ ™ v̌|. The language user now has access to both of these candidates via the rank-ordering that EVAL imposes on the candidate set. However, the likelihood that the language user will actually access a candidate depends on the position that the candidate occupies in the rank-ordering. The higher a candidate appears in the rank-ordering, the more likely it is that the language user will access it as output. Since the deletion candidate appears higher than the retention candidate, the prediction is that the language user will be more likely to select the deletion candidate as output. We therefore expect to see the deletion variant more frequently than the retention variant.

Now consider tri-syllabic forms. When the final unstressed vowel is deleted from a tri-syllabic form, a bi-syllabic form with an unstressed vowel in the final syllable results. To make this concrete, consider the form *pazinu* ‘I knew’. This form has two pronunciations: the faithful retention form [(pá.zi)nu] and the unfaithful deletion form [(pá.zin)]. Both of these forms have a vowel in a final unstressed syllable, and both of

them therefore violate *v̌]$_{\sigma}$]$_{\omega}$. It cannot be this constraint that drives deletion in tri-syllabic forms. However, there is another difference between retention and the deletion forms. The retention variant has a final syllable that is not parsed into a foot and therefore earns a violation in terms of the constraint PARSE-σ that requires every syllable to be parsed into a foot. I claim that it is PARSE-σ that drives deletion in tri-syllables. Since tri-syllables also show more deletion than retention, we need the deletion candidate to be rated better than the retention candidate by EVAL. This can be achieved if the highest ranked constraint that disfavors retention outranks the highest ranked constraint that disfavors deletion – i.e. we need the ranking ||PARSE-σ o MAX||.[8] This is illustrated in the tableau in (6).

(6) **Deletion preferred over retention in tri-syllabic forms**

			PARSE-σ	*v̌]$_{\sigma}$]$_{\omega}$	MAX
/pazinu/	2	(pá.zi)nu	*	*	
	1	(pá.zin)		*	*

Output of EVAL

(pá.zin) MAX
|
(pá.zi)nu PARSE-σ

Both the retention and the deletion candidate violate *v̌]$_{\sigma}$]$_{\omega}$ and this constraint therefore does not distinguish between the two candidates. However, the retention candidate violates PARSE-σ while the deletion candidate violates MAX. Because of the ranking ||PARSE-σ o MAX|| EVAL rates the deletion candidate better than the retention candidate – i.e. |∅ ™ v̌|. The deletion candidate consequently occupies a higher slot in the rank-ordering that EVAL imposes on the candidate set, is more accessible as output, and is predicted to be the more frequently observed variant.

[8] See the discussion in §2.4 for more discussion about the motivation for this ranking.

This analysis accounts for the following two facts: (i) that variation is observed – by allowing language users access to candidates beneath the best candidate; (ii) that deletion is more frequent than retention – since the deletion candidate occupies a higher slot in the rank-ordered candidate set and is therefore more accessible.

2.2 Inter-contextual variation: more deletion in tri-syllabic than in bi-syllabic words

Even though we can now account for the fact that deletion is observed more often than retention in both bi- and tri-syllabic forms, we still have to account for the fact that tri-syllabic words are associated with a higher deletion rate than bi-syllabic words. To account for this, I call on the ability of EVAL to make comparisons between morphologically unrelated forms. The non-deletion candidate of a tri-syllabic word is more marked than the non-deletion candidate of a bi-syllabic word, and the drive to delete is therefore stronger in tri-syllabic words. In order to account for this we have to consider the non-generated comparison set that contains the faithful candidates from bi-syllabic and tri-syllabic inputs.

The fully faithful candidates from tri-syllabic and bi-syllabic inputs have the following form: [(σ́.σ)σ] and [(σ́.σ)] – cf. [(pá.zi)nu] and [(tá.ka)]. Both of these forms violate $*\check{v}]_{\sigma}]_{\omega}$, and this constraint can therefore not distinguish between them. However, the tri-syllabic form also violates the constraint PARSE-σ that requires all syllables to be parsed into feet. Because of its final unfooted syllable, the faithful candidate of a tri-syllabic word is more marked than the faithful candidate of a bi-syllabic word. This explains why tri-syllabic words are associated with higher deletion rates – a faithful tri-syllabic word is more marked than a faithful bi-syllabic word, and therefore the drive to delete is stronger in tri-syllabic words. This is shown in the tableau in (7).

(7) **Non-generated comparison set containing the faithful candidates of a bi-syllabic and a tri-syllabic input**

		PARSE-σ	*v̆]$_{\sigma}$]$_{\omega}$	MAX
1	(tá.ka)		*	
2	(pá.zi)nu	*	*	

Output of EVAL

(tá.ka)

|

(pá.zi)nu$_{\text{PARSE-}\sigma}$

By allowing EVAL to compare forms that are not related to each other via a shared input, we can formally capture the intuition that retention of the unstressed vowel is more marked in tri-syllabic forms than in bi-syllabic forms.

2.3 Limits on variation

Up to this point in the discussion I have considered only two candidates for each input – namely a faithful retention candidate and an unfaithful deletion candidate. However, the generated candidate sets undoubtedly contain many more candidates than just these two. Under the rank-ordering model of EVAL, each of these candidates will occupy a slot in the rank-ordering. And under the assumption that the language user has access to levels below the highest level in this ordering, we predict that these other candidates should also be observed as variants. This problem becomes particularly acute when we consider candidates other than the deletion candidate that also avoid violation of the deletion inducing constraints *v̆]$_{\sigma}$]$_{\omega}$ and PARSE-σ. For the purposes of the illustration here I will discuss only one example of this kind of candidate, namely candidates that impose an alternative foot structure on the output.

I will explain in two steps how the non-observed forms are ruled out. First, I will show that these forms are less well-formed than the two variants that are observed (the retention and the deletion candidate). The non-observed forms will therefore occupy a

lower slot in the rank-ordering that EVAL imposes on the candidate set. Even had they been observed as variants, the prediction would then be that they will be observed less frequently than the retention and the deletion candidates. After this has been established, I will introduce the notion of the "critical cut-off". The critical cut-off is a position on the constraint hierarchy that represents a line that is crossed only under great duress – only when no other options are available will the language user be willing to access candidates eliminated by constraints ranked higher than the cut-off. I will then argue that the non-observed forms are all eliminated by constraints ranked higher than the cut-off, and that this explains why they are never accessed as variant outputs.

(8) **Other candidates that avoid a *v̆]$_{\sigma}$]$_{\omega}$-violation**

Input	**Observed**	**Not observed**
/taka/	[(ták)]	*[(tá)(kà)]
/pazinu/	[(pá.zin)]	*[(pá.zi)(nù)]

First, let us consider why the un-observed forms from the right hand column in (8) are less well-formed than the faithful retention candidate and the deletion candidate. As I mentioned in §1, Latvian feet are syllabic trochees. This means that Latvian feet are preferably bi-syllabic. The preference for bi-syllabic feet can be enforced by the constraint FTBIN-σ which is violated by any mono-syllabic (degenerate) foot. The un-observed forms in the right hand column of (8) all have such degenerate feet and therefore all violate FTBIN-σ. Initially it therefore seems that this constraint could account for the fact that these forms are less well-formed than the forms that are actually attested. However, there is a complication – Latvian does tolerate degenerate feet. There

are many mono-syllabic words in Latvian, and these words are not augmented by epenthesis in order to create bi-syllabic feet. Also, when a vowel is deleted from a bi-syllabic word, it is usually the case that a mono-syllabic form is created. And vowels delete very often from bi-syllabic forms. In fact, one of the examples in (8) shows this – the form /taka/ can be pronounced as [(ták)]. Latvian does therefore tolerate degenerate mono-syllabic feet.

However, Latvian tolerates mono-syllabic feet only in strong positions – i.e. only if the foot happens to be the foot that carries the main stress of the prosodic word, can it be a mono-syllabic foot. The result is that the only mono-syllabic feet tolerated in Latvian are in mono-syllabic prosodic words. The single syllable in these forms is then parsed into a foot, which, since it is the only foot, is also the main foot and therefore receives main stress. No foot that receives only secondary stress can ever be mono-syllabic. This is a very well documented cross-linguistic pattern. In fact, Hayes (1995:87) claims that languages can be divided into two classes – those that do not allow degenerate feet al all, and those that allow degenerate feet only in strong position (i.e. if the foot is also the main foot of a prosodic word). Based on this fact, I propose that there are actually two versions of the foot binarity constraint – one that applies to all feet and one that applies only to feet that do not function as the head of some prosodic word. Latvian tolerates violation of the general constraint – mono-syllabic words are tolerated because in these forms the single syllable is parsed into the head foot of the prosodic word. Latvian does not tolerate violation of the constraint against non-head degenerate feet – this is why a form with an uneven number of syllables always ends in an unfooted syllable.

(9) **Foot binarity constraints**

FTBIN-σ

Feet are bi-syllabic.

FTBIN-$\sigma_{\text{Non-main foot}}$

A foot that is not the head of a prosodic word is bi-syllabic.

We now have to figure out where these two foot binarity constraints should be ranked. Let us start with the general constraint FTBIN-σ. We can show that this constraint must rank below $*\check{v}]_{\sigma}]_{\omega}$, i.e. $||*\check{v}]_{\sigma}]_{\omega}$ o FTBIN-σ||. A bi-syllabic input such as /taka/ can be pronounced with or without its final vowel, i.e. as [(tá.ka)] or as [(ták)]. The retention candidate [(tá.ka)] violates only $*\check{v}]_{\sigma}]_{\omega}$, while the deletion candidate [(ták)] violates both MAX and FTBIN-σ. Of these two variant pronunciations the deletion candidate is the more frequently observed variant. This means that EVAL should rate this candidate better than the retention candidate. This, in turn, is only possible if the highest ranked constraint that disfavors the retention candidate is ranked higher than the highest ranked constraint that disfavors the deletion candidate. This means that we need the ranking $||*\check{v}]_{\sigma}]_{\omega}$ o {MAX, FTBIN-σ}||. I have already argued for the ranking $||*\check{v}]_{\sigma}]_{\omega}$ o MAX|| in §2.1 above. The only new ranking that we need to add is that between $*\check{v}]_{\sigma}]_{\omega}$ and FTBIN-σ. The need for this ranking is shown in the tableau in (10).[9]

---

[9] There is no evidence for the ranking ||FTBIN-σ o MAX||. However, throughout this dissertation I follow the principle of "ranking conservatism" (Itô and Mester, 1999, 2003, Tesar and Smolensky, 1998, 2000). This principle is based on the assumption that the original state of the grammar has the ranking ||Markedness o Faithfulness|| (Smolensky, 1996). The claim is that this ranking should be preserved unless there is evidence to contrary. I will therefore rank faithfulness constraints as low as possible, promoting them to above a markedness constraint only if there is positive evidence for this.

(10) **||*ˇv]$_\sigma$]$_\omega$ o FTBIN-σ||**

			*ˇv]$_\sigma$]$_\omega$	FTBIN-σ	MAX
/taka/	1	(ták)		*	*
	2	(tá.ka)	*		

Output of EVAL

(ták) $_{\text{FTBIN-}\sigma}$

|

(tá.ka) $_{*\check{v}]\sigma]\omega}$

The faithful candidate violates *ˇv]$_\sigma$]$_\omega$, and the deletion candidate violates FTBIN-σ. However, because of the ranking ||*ˇv]$_\sigma$]$_\omega$ o FTBIN-σ|| EVAL rates the deletion candidate better, and imposes the following ordering on the candidate set: |∅ ™ ˇv|. The deletion candidate therefore occupies a higher slot in the rank-ordering and is more accessible. From this follows that it would be observed more frequently as output.

Now we can consider where the constraint FTBIN-$\sigma_{\text{Non-main foot}}$ should be ranked. This time consider a tri-syllabic input such as /pazinu/. One of the variant pronunciations for this input is the faithful preservation candidate [(pá.zi)nu] that violates *ˇv]$_\sigma$]$_\omega$ and PARSE-σ. The candidate that imposes an alternative foot structure on the output, i.e. [(pá.zi)(nù)], is not observed. Of the constraints that we are dealing with here, this non-observed form violates only FTBIN-$\sigma_{\text{Non-main foot}}$. If we think in terms of variation about the fact that [(pá.zi)(nù)] is never observed, we can say that [(pá.zi)(nù)] is observed less frequently than the variants [(pá.zi)nu] – in fact [(pá.zi)(nù)] is never observed while [(pá.zi)nu] is sometimes observed. This would require that [(pá.zi)(nù)] occupies a lower slot in the rank-ordering that EVAL imposes on the candidate set than [(pá.zi)nu]. This can be achieved if the highest ranked constraint that disfavors [(pá.zi)(nù)] outranks the highest ranked constraint that disfavors [(pá.zi)nu] – i.e. we need FTBIN-$\sigma_{\text{Non-main foot}}$ to outrank *ˇv]$_\sigma$]$_\omega$, and PARSE-σ. This is shown in the tableau in (11).

(11) **||FTBIN-σ$_{\text{Non-main foot}}$ o {PARSE-σ, *ˇv]σ]ω}||**

			FTBIN-σ Non-main foot	PARSE-σ	*ˇv]σ]ω
/pazinu/	1	(pá.zi)nu		*	*
	2	(pá.zi)(nù)	*		

Output of EVAL

(pá.zi)nu {PARSE-σ, *ˇv]σ]ω}

|

(pá.zi)(nù) FTBIN-σNon-main foot

We now have the ranking ||FTBIN-σ$_{\text{Non-main foot}}$ o {PARSE-σ, *ˇv]σ]ω} o FTBIN-σ o MAX||. With this ranking we can show that the non-observed forms (with degenerate non-head feet) are less well-formed than the observed variants. The non-observed forms do occupy a slot in the rank-ordering, but they will always occupy a lower slot. This is shown in the tableaux in (12).

(12) a. **Bi-syllabic: [(tá)(kà)] less well-formed than [(ták)] and [(tá.ka)]**

			FTBIN-σ Non-main foot	PARSE-σ	*ˇv]σ]ω	FTBIN-σ	MAX
/taka/	2	(tá.ka)			*		
	1	(ták)				*	*
	3	(tá)(kà)	*			**	

Output of EVAL

(ták) FTBIN-σ

|

(tá.ka) *ˇv]σ]ω

|

(tá)(kà) FTBIN-σNon-main foot

b. **Tri-syllabic: [(pá.zi)(nù)] less well-formed than [(pá.zin)] and [(pá.zi)nu]**

			FTBIN-σ Non-main foot	PARSE-σ	*ˇv]σ]ω	FTBIN-σ	MAX
/pazinu/	2	(pá.zi)nu		*	*		
	1	(pá.zin)			*		*
	3	(pá.zi)(nù)	*			*	

Output of EVAL

(pá.zin) *ˇv]σ]ω

|

(pá.zi)nu PARSE-σ

|

(pá.zi)(nù) FTBIN-σNon-main foot

These tableaux in (12) show that the non-observed candidates [(tá)(kà)] and [(pá.zi)(nù)] do occupy a slot in the rank-ordering that EVAL imposes on the candidate set. However, since these two candidates violate the constraint FTBIN-$\sigma_{\text{Non-main foot}}$ that is ranked higher than any of the constraints violated by the observed variants, the non-observed candidates occupy a lower slot in the rank-ordering. This completes the first step in explaining how to limit variation to only the observed variants – I have now shown that the non-observed forms are less well-formed than all of the observed forms. However, this still does not explain why these forms are never observed as variants. Under the assumption that the language user has potential access to the full candidate set via the rank-ordering that EVAL imposes on the candidate set, these non-observed forms should also be accessible. The best we can say at this moment is that these forms should be observed as less frequent variants. We cannot yet explain that they are never selected as output. In order to explain this, I introduce the concept of the critical cut-off.

It seems to be the case that there are certain constraints that Latvian are willing to violate – these are the constraints violated by the actually observed variants, i.e. MAX, *v̌]$_{\sigma}$]$_{\omega}$, PARSE-σ, and FTBIN-σ. However, there are certain other constraints of which Latvian will not tolerate violation. One such a constraint is the constraint FTBIN-$\sigma_{\text{Non-main foot}}$ violated by the non-observed forms [(tá)(kà)] and [(pá.zi)(nù)]. The constraint set can therefore be divided into the set of constraints that the language is willing to violate in a variable phenomenon, and the set of constraints that the language is not willing to violate in a variable phenomenon. I claim there is a critical cut-off on the constraint hierarchy that distinguishes these two classes of constraints from each other. Only those constraints that a language is willing to violate in a variable phenomenon are

ranked lower than the critical cut-off. All others are ranked higher than the critical cutoff. In Latvian all of constraints violated by the observed variants rank below the cut-off (i.e. MAX, *$\check{v}]_{\sigma}]_{\omega}$, PARSE-σ, and FTBIN-σ), while the constraints violated by the non-observed forms (FTBIN-$\sigma_{\text{Non-main foot}}$) rank above the cut-off. The role of the critical cut-off is important enough that I state it explicitly in (13).

(13) **Critical cut-off**[10]

a. The critical cut-off is a point on the constraint hierarchy.

b. In general the language user will not select as output candidates that are eliminated by constraints ranked higher than the critical cut-off – i.e. if there is any candidate that does not violate any constraint ranked higher than the cut-off, then no candidate that does violate a constraint ranked higher than the cut-off will be selected as output.

In Latvian the critical cut-off appears right between FTBIN-$\sigma_{\text{Non-main foot}}$ and the two markedness constraints PARSE-σ and *$\check{v}]_{\sigma}]_{\omega}$. In (14) I repeat the tableaux from (12) above, this time with the critical cut-off included. In both of these tableaux the non-observed forms do occupy a slot in the rank-ordering that EVAL imposes on the candidate set. However, these forms are disfavored by FTBIN-$\sigma_{\text{Non-main foot}}$ which ranks higher than the cut-off. Since there are candidates available that are not disfavored by any constraint ranked above the cut-off, these candidates will never be accessed as variable outputs. (On the typographical conventions used in these tableau, see Chapter 1 §2.2.3.)

---

[10] For more general discussion about the critical cut-off, see Chapter 1 §2.2.3.

(14) a. **Bi-syllabic: With the critical cut-off included**

			FtBin-σ Non-main foot	Parse-σ	*v̌]σ]ω	FtBin-σ	Max
/taka/	2	(tá.ka)			*		
	1	(ták)				*	*
	3	(tá)(kà)	*!			**	

Output of EVAL

(ták) FtBin-σ

(tá.ka) *v̌]σ]ω

Cut-off

(tá)(kà) FtBin-σNon-main foot

b. **Tri-syllabic: with the critical cut-of included**

			FtBin-σ Non-main foot	Parse-σ	*v̌]σ]ω	FtBin-σ	Max
/pazinu/	2	(pá.zi)nu		*	*		
	1	(pá.zin)			*		*
	3	(pá.zi)(nù)	*!			*	

Output of EVAL

(pá.zin) *v̌]σ]ω

(pá.zi)nu Parse-σ

Cut-off

(pá.zi)(nù) FtBin-σNon-main foot

By introduction of the critical cut-off point we are able to account for the fact that variation is usually limited to only the best two or three candidates. Although all candidates occupy a slot in the rank-ordering that EVAL imposes on the candidate set, language users will usually not access this rank-ordering to an arbitrary depth. Candidates eliminated by constraints ranked higher than the cut-off are not accessed as variable outputs. In the discussion in this section I have considered only one kind of non-observed form. However, all other non-observed forms will be treated in exactly the same manner – each of them has to violate at least one constraint ranked higher than the cut-off.[11]

[11] Although language users do not access candidates eliminated by constraints ranked higher than the cut-off in variable phenomena, these candidates are in principle accessible to language users. See the discussion in Chapter 1 §2.2.3 for how language users can access these candidates in categorical

2.4 Ranking arguments

In classic OT, constraint rankings are motivated by "ranking arguments" (McCarthy, 2002b: 4-5, 30-39). A ranking argument is constructed to show that the ranking between two constraints is crucial for the selection of the single optimal candidate – i.e. the opposite ranking will result in selection of an incorrect optimal candidate. In the rank-ordering model of EVAL we have to think differently about ranking arguments. Ranking arguments can now also be frequency based. In this section I will point out how frequency based ranking arguments can be constructed.

Consider again tableau (5) from above in which I have relied in the ranking ||*v̌]$_{\sigma}$]$_{\omega}$ o MAX||. I repeat this tableau in (15). However this time I include the location of the critical cut-off. In (14) just above I have shown that both *v̌]$_{\sigma}$]$_{\omega}$ and MAX rank below the critical cut-off. In this tableau C stands for any constraint ranked higher than the cut-off.

(15) **[(tá.ka)] and [(ták)] as only variants under ||*v̌]$_{\sigma}$]$_{\omega}$ o MAX||**

			C	*v̌]$_{\sigma}$]$_{\omega}$	MAX
/taka/	2	(tá.ka)		*	
	1	(ták)			*
	3	All other cands	*!		

Output of EVAL

(ták) MAX

|

(tá.ka) *v̌]σ]ω

┼ Cut-off

Other cands C

For the input /taka/ there are two possible outputs, namely [(tá.ka)] and [(ták)]. As the tableau in (15) shows, these are indeed the only two candidates that are selected as

phenomena. See also Chapter 6 for evidence that language users do access information about these candidates in word-likeness judgments and in lexical decision tasks.

output with the ranking ||*ˇ]$_{\sigma}$]$_{\omega}$ o MAX||. However, this ranking is not crucial to assure that these two are indeed the only two observed outputs. Even under the opposite ranking only these two will be selected. This is shown in the tableau in (16).

(16) **Even under opposite ranking ||MAX o *ˇ]$_{\sigma}$]$_{\omega}$|| [(tá.ka)] and [(ták)] are the only variants**

			C	MAX	*ˇ]$_{\sigma}$]$_{\omega}$
/taka/	1	(tá.ka)			*
	2	(ták)		*	
	3	All other cands	*!		

Output of EVAL

(tá.ka) *ˇ]σ]ω

(ták) MAX

Cut-off

Other cands C

Even in this tableau only the two actually observed variants are predicted as possible. However, under this ranking it is predicted that the faithful, non-deletion candidate should be the more frequently observed variant. And we know that this is not true. The argument for the ranking ||*ˇ]$_{\sigma}$]$_{\omega}$ o MAX|| can therefore not be based on a ranking argument of classic OT. It is not the case that the opposite ranking selects a non-observed form as output. This argument rather takes the following form: Under the opposite ranking the variant that occurs more frequently is wrongly predicted to occur less frequently.

Classic OT ranking arguments can still be used in categorical processes – i.e. where there is indeed only one correct output for some input. In such situations we can show that some rankings will lead to selection of the wrong output. However, even in these cases it is not necessary to use a classic OT ranking argument. We can think even about these ranking arguments in terms of frequency. To see how this can be done, let us

consider an example where a classic OT ranking argument can be used. Mono-syllabic words in Latvian are not augmented by epenthesis. This means that the output of a mono-syllabic word will violate the constraint FTBIN-σ, requiring feet to be bi-syllabic. Latvian could have avoided this violation by augmenting the word via epenthesis, i.e. at the expense of a DEP-violation. The fact that Latvian does not do this is evidence for the ranking ||DEP ο FTBIN$_{\sigma}$||. A classic OT ranking argument can be constructed for this ranking since it can be shown that under the opposite ranking the incorrect (augmentation) candidate will be selected as optimal. This is shown in (17).

(17) **Ranking argument for ||DEP ο FTBIN-σ||:**
With the opposite ranking the incorrect output candidate is selected

/nest/ "to carry"	FTBIN-σ	DEP
a. (nést)	*!	
b. (nés.tV)		*

Under the ranking ||FTBIN-σ ο DEP||, it is predicted that the augmentation candidate is better than the faithful candidate. However, we know that Latvian does not augment mono-syllabic inputs to avoid a violation of FTBIN-σ. This gives us the evidence that DEP outranks FTBIN-σ.

We can view a categorical phenomenon such as this in terms of variation. We can consider both the faithful candidate [(nést)] and the augmentation candidate [(nés.tV)] as variants. But this is an extreme example of variation where one of the variants occurs infinitely more frequently than the other. The faithful [(nést)] occurs 100% of the time while the augmentation candidate [(nés.tV)] occurs 0% of the time. Now the ranking argument in (17) can also be recast in terms of frequency. Under the ranking ||FTBIN-σ ο

DEP|| the variant that is observed more frequently (in fact 100% of the time) is predicted to occur less frequently. Therefore, we need the ranking ||DEP o FTBIN-σ|| between these two constraints.

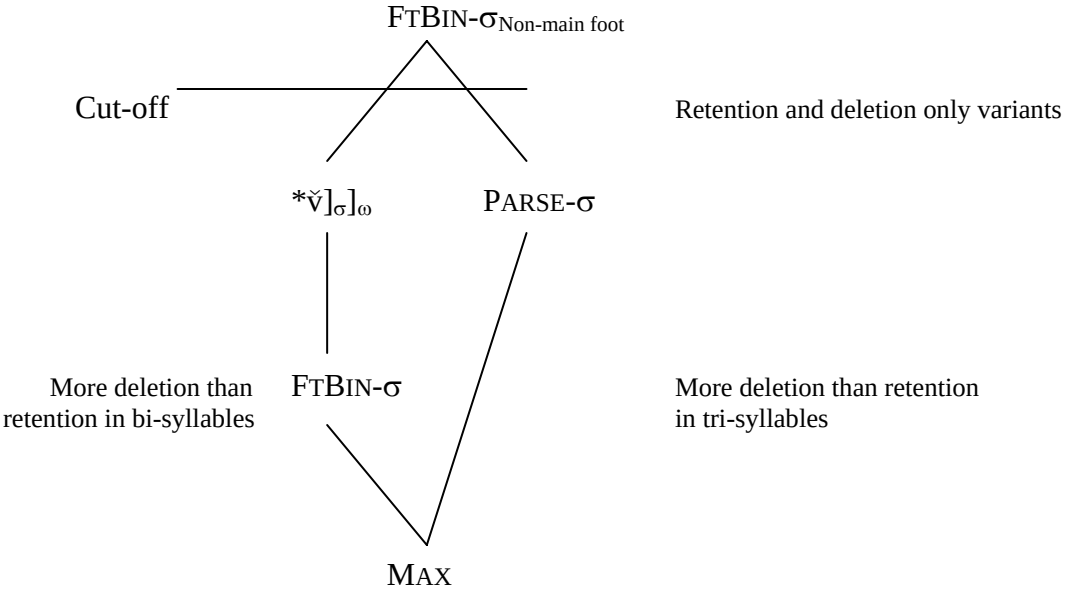
2.5 Summary

In (18) I list all of the rankings that I have argued for above. The first column contains the ranking, the second column a short motivation for the particular ranking, and the last column a reference to where in the preceding text that particular ranking was discussed. After the table I give a graphic representation of the rankings.

(18) **Summary of rankings necessary to explain vowel deletion in Latvian**

Ranking	**Motivation**	**Where?**
$*\check{v}]_{\sigma}]_{\omega}$ o MAX	More deletion than retention in bi-syllabic forms	(5)
PARSE-σ o MAX	More deletion than retention in tri-syllabic forms	(6)
$*\check{v}]_{\sigma}]_{\omega}$ o FTBIN-σ	More deletion than retention in bi-syllabic forms	(10)
FTBIN-$\sigma_{\text{Non-main foot}}$ o Cut-off	No degenerate feet that are not the head of a prosodic word	(14)
Cut-off o MAX, $*\check{v}]_{\sigma}]_{\omega}$, PARSE-σ, FTBIN-σ	Variation between deletion and retention observed	(14)
FTBIN-σ o MAX	Ranking conservatism	(10), footnote 9

(19) **Graphic representation of the rankings for Latvian vowel deletion**



CHAPTER 4

VOWEL DELETION IN FAIALENSE PORTUGUESE

In European Portuguese unstressed vowels variably delete. The word *bato* 'I beat' can therefore be pronounced as either [bátu] or [bát]. This deletion pattern is not random, but is strongly influenced by phonological considerations – certain prosodic contexts are associated with higher deletion rates than others, certain vowels delete more often than others, etc. (Mateus and d'Andrade, 2000:18, Silva, 1997, 1998). In this chapter I provide an analysis of this variable deletion pattern within the rank-ordering model of EVAL. In the discussion below I will assume familiarity with the rank-ordering model of EVAL and with how this model deals with variation. For a general discussion on this model, see Chapters 1 and 3.

European Portuguese, like many languages, allows only a subset of its full vowel inventory in unstressed syllables. The smaller vowel inventory observed in unstressed positions is achieved through vowel reduction. This vowel reduction process is categorical – that is, a particular input vowel either does or does not reduce. In addition to reduction, European Portuguese also has a variable process that deletes vowels from unstressed syllables.[1] The result of this is that many European Portuguese words have

[1] Brazilian Portuguese also has a reduced vowel inventory in unstressed syllables (Fails and Clegg, 1992, Mateus and d'Andrade, 2000:17-18, Thomas, 1974:4-7). However, unlike European Portuguese, Brazilian Portuguese does not apply vowel deletion (Mateus and d'Andrade, 2000:46, 134-135, Oliviera, 1993:9). Silva (1997:307, endnote 2) attributes this to the fact that European Portuguese is a "stress timed" language, while Brazilian Portuguese is syllable timed. (See also Parkinson (1988:141-142) for a classification of Brazilian Portuguese as syllable timed. However, see Major (1981, 1985, 1992) for some arguments to the contrary.) Stressed timed languages are much more likely to have vowel reduction and/or vowel deletion processes than syllable timed languages.

two possible pronunciations: One with a vowel in the unstressed syllable,[2] and another in which the vowel from the unstressed syllable has been deleted.

(1) **Examples of variation in European Portuguese**

/sɛlo/	[sɛ́lu] ~ [sɛ́l]	'stamp'
/idade/	[idádi] ~ [idád]	'age'
/muʎɛɾ/	[muʎɛ́ɾ] ~ [mʎɛ́ɾ]	'woman'

In the rest of this chapter I offer a detailed account of this variation pattern within the rank-ordering model of EVAL. Since there is no quantitative data available on vowel deletion in standard (Lisbon) European Portuguese, I will use data on vowel deletion in an Azorean dialect of Portuguese, namely that spoken on the island of Faial. Although there are differences between standard European Portuguese and Azorean Portuguese, the basic patterns observed in the vocalic phonology of these two varieties of European Portuguese are very similar.[3] In the discussion below, all references to Portuguese should

---

[2] This vowel can be different from its input correspondent if the input vowel is one of the vowels subject to reduction. However, it can also be identical to the input vowel if the input vowel is not one of the vowels subject to reduction.

[3] As far as possible I will rely on sources specifically about the phonology of Azorean Portuguese (Rogers, 1948, 1949, Silva, 1988, 1997, 1998). However, these sources do not always present us with enough information. The vocalic phonology of Azorean Portuguese is sufficiently similar to that of standard European Portuguese that it justifies the use of grammars on standard European Portuguese where the information on Azorean Portuguese is insufficient – see for instance Mateus and d'Andrade (2000:2) who claim that the phonological differences between dialects of European Portuguese occur mainly "in the fricative consonant system" and not in the vocalic system. Mateus and d'Andrade also claim that "the dialects on … the Azores, while they have their own peculiarities, share the general characteristics of the central-southern dialects" (2000:2). Lisbon Portuguese is a "central-southern" dialect. This provides more motivation for the use of grammars of standard European Portuguese to supplement information on Azorean Portuguese.

Even more motivation for using information on standard European Portuguese to supplement that on Faialense Portuguese comes from Rogers (1949:48): "It has been shown that the pronunciation of Portuguese on the Madeiran and eastern Azorean islands is quite different from that of the standard language heard on the European continent. This divergence from the standard language does not hold for the central and western Azores." Faial is a central Azorean island.

therefore be interpreted as referring equally to standard European Portuguese and Faialense Portuguese, unless otherwise stated.

The rest of this chapter is structured as follows. In §1 I will discuss the basics of the vocalic phonology of Faialense Portuguese, focusing on the processes that apply in unstressed syllables. In §2 I will present an OT account of the vowel reduction in unstressed syllables. Finally, in §3 and §4, I will give an OT account of variable vowel deletion, and show how this process interacts with vowel reduction. I postpone a discussion of alternative accounts of variation until the end of the section of the dissertation that deals with variable phenomena (Chapter 5 §3).

1. The basic vocalic phonology of Faialense Portuguese

1.1 The oral vowels of Faialense Portuguese

In (2) I give the vowel inventory of standard European Portuguese, which according to Silva (1997:298) "is also found in most Azorean varieties of the language, including that spoken on the island of Faial". The diphthongs are immune to the vowel lenition processes, and nasal vowels undergo a different kind of reduction than oral vowels (Silva, 1997:299).[4] I will limit myself here to only the oral monophthongs. The features that I will assume for the vowels in the rest of this discussion are given in (3). The table is followed by some discussion of the features.[5]

---

[4] The nasal vowel inventory is already smaller than the oral vowel inventory. In post-tonic position, nasal vowels do reduce. However, in addition to "reduction" they also undergo a process of diphthongization in post-tonic position.

[5] The vowel inventory given here, agrees with that proposed by Silva (1997:298) and Parkinson (1988). Mateus and d'Andrade (2000:17-18) agree on the inventories for the stressed vowels. However, they

(2) **Oral vowel inventory of Portuguese**

a. Stressed syllables

i u

e o

ɛ ɔ

ɐ

a

b. Unstressed syllables

i u

ə

ɐ

(3) **The features of the Faialense Portuguese vowels**

	i	u	e	o	ɛ	ɔ	ɐ	a	ə
High	+	+	-	-	-	-	-	-	-
Low	-	-	-	-	-	-	+	+	-
Front	+	-	+	-	+	-	-	-	-
Back	-	+	-	+	-	+	-	-	-
ATR	+	+	+	+	-	-	+	-	-

Height. I am making a three level height distinction. In particular, the vowels can be ordered from low to high as follows: [a, ɐ] > [e, o, ɛ, ɔ, ə] > [i, u]. The two low vowels [a, ɐ], the two mid front vowels [e, ɛ], and two mid back vowels [o, ɔ] are not formally distinguished from each other by a height feature, but rather by the feature ATR. In absolute acoustic terms the [+ATR] (or tense) vowels are all higher than their [-ATR] (or lax) counterparts. This is confirmed by a spectrographic analysis of the tonic vowels of European Portuguese performed by Martins (1964 and 1973). The average F_1 values that

do not recognize schwa in unstressed syllables. The vowels that Silva indicates as [ə] are considered to be [ɨ] by Mateus and d'Andrade. I follow Silva and Parkinson here, primarily because the data on vowel deletion in Faialense Portuguese are reported by Silva, and it is therefore easier to interpret his data if I assume same the vowel inventory that he assumes.

This difference between authors on the vowel inventory is not surprising. The high non-back unrounded vowels and schwa are not only acoustically quite similar, but they also do sometimes pattern together phonologically. For some discussion on the close phonological relationship between high non-back unrounded vowels and schwa in Tiberian Hebrew, see Coetzee (1996a, 1996b, 1999a:122-126) and Garr (1989).

Martins found for these sets of vowels are given in (4) below. An advanced tongue root is known to raise a vowel slightly. The fact that the [+ATR] vowels are higher than their [-ATR] counterparts is therefore not surprising.

(4) **F_1 values for low and mid vowels in European Portuguese**[6]

	F_1		F_1		F_1
e	403	o	426	ɐ	511
ɛ	501	ɔ	531	a	626

Frontness/Backness. I distinguish front [e, ɛ, i], central [a, ɐ, ə], and back vowels [o, ɔ, u]. This deviates from Mateus and d'Andrade (2000:30) who distinguish only front and back vowels, and who classify [a, ɐ] as back. However, it agrees with Silva (1988, 1997, 1998), Fails and Clegg (1992), and Parkinson (1988:132). The classification of [a, ɐ, ə] as central is also confirmed by the Martins's spectrographic analysis. The average F_2 values for the vowels of European Portuguese in the table in (5) are from *Figure 12* in her paper (p. 312). Since Martins investigated only vowels in stressed syllables, she does not report values for schwa, which occurs only in unstressed syllables in Portuguese.

---

6 Martins's study was done on standard peninsular Portuguese and not Azorean Portuguese. However, there is no reason to assume that Azorean and peninsular Portuguese would differ in the basic pattern. Additionally, even Brazilian Portuguese has the same F_1 relationships between these vowels. The F_1 values for Brazilian Portuguese vowels below are from Fails and Clegg (1992:36). Brazilian Portuguese does not have the vowel [ɐ] in tonic position, and comparison between [a] and [ɐ] is therefore not possible.

	F_1		F_1
e	383	o	399
ɛ	539	ɔ	545

(5) **F_2 values for low and mid vowels in European Portuguese**[7]

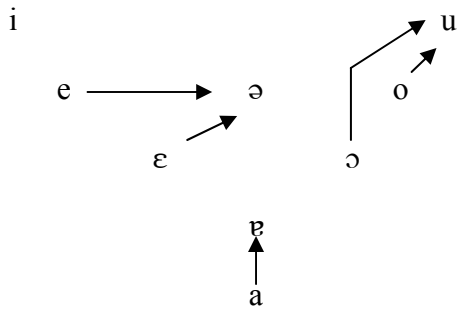
	F_2		F_2		F_2
i	2344	ɐ	1602	u	678
e	2084	a	1326	o	864
ɛ	1893			ɔ	994

Roundness. Roundness is completely redundant – all and only the back vowels are round. I therefore do not include a separate feature [round].

1.2 Vowel reduction and deletion in unstressed syllables

The schematic representation of the reduction process in (6) is adapted from Silva (1997:299, 1998:175).[8]

(6) **Vowel reduction in unstressed syllables**



[7] Again, Brazilian Portuguese shows the same pattern (Fails and Clegg, 1992:36).

	F_2		F_2		F_2
i	2149	a	1264	u	896
e	1936			o	780
ɛ	1659			ɔ	939

[8] Silva (1997:299, 1998:175) and Mateus and d'Andrade (2000:18) claim that /i/ reduces to schwa in post-tonic position. However, the stress placement rules of Portuguese are such that /i/ will never be parsed into a post-tonic position. Portuguese words are usually stressed on the penultimate syllable. However, when the final syllable contains one of the two high vowels /i, u/, stress is attracted to this final syllable (Silva, 1997:299, Thomas, 1974:3). The result is that it is not possible for an /i/ to occur in post-tonic position. The statement that post-tonic /i/ reduces is therefore a vacuous statement. Silva has also confirmed this to me in personal communication.

(7) **Examples of the reduction in unstressed syllables**[9]

/e/ → [ə]:	*selo*	[sé̱lu] "stamp"	vs.	*selar*	[sə̱láɾ] "to stamp"
/ɛ/ → [ə]:	*selo*	[sɛ̱́lu] "I stamp"	vs.	*selar*	[sə̱láɾ] "to stamp"
/a/ → [ɐ]:	*paga*	[pá̱gɐ] "s/he pays"	vs.	*pagar*	[pɐ̱gáɾ] "to pay"
/ɔ/ → [u]:	*forço*	[fɔ̱́ɾsu] "I oblige"	vs.	*forçar*	[fu̱ɾsáɾ] "to oblige"
/o/ → [u]:	*força*	[fó̱ɾsa] "strength"	vs.	*forçar*	[fu̱ɾsáɾ] "to oblige"

(Mateus and d'Andrade, 2000:17, 20)

Crosswhite (2000a, 2000b, 2001) distinguishes two types of vowel reduction, namely *contrast enhancing* reduction and *prominence neutralizing* reduction. Contrast enhancing reduction is characterized by the avoidance of non-peripheral vowels in perceptually weak positions such as unstressed syllables. In these kinds of systems, the inventory of reduced vowels therefore often consists of the three peripheral vowels [i, u, a]. Prominence neutralizing vowel reduction is characterized by a drive to have elements with similar prominence characteristics co-occur. Vowels of lower sonority are less prominent than vowels of higher sonority, and unstressed syllables are less prominent than stressed syllables. These kinds of reduction processes therefore usually result in the replacement of high sonority vowels with lower sonority vowels in unstressed syllables.

The reduction process observed in Portuguese is prominence neutralizing reduction. Each of the reductions in unstressed syllables replaces a vowel of higher sonority with a vowel of lower sonority. This rests on two assumptions about the sonority

---

[9] In these examples I have replaced the [ɨ] of Mateus and d'Andrade with [ə]. For more on this, see footnote 5 above.

of vowels: (i) Schwa is the least sonorous vowel.[10] (ii) Lower vowels are more sonorous than higher vowels (Parker, 2002). Under these assumptions the Portuguese vowels can be ordered according to their sonority as in (8).

(8) **Sonority scale for Portuguese vowels**

a > ɐ > {ɛ, e, ɔ, o}[11] > u > i > ə

Aside from this positive evidence that Portuguese vowel reduction is of the prominence reduction kind, there is also negative evidence. If Portuguese vowel reduction was a contrast enhancing process, then two of the reduction mappings could not be explained. (i) The corner vowel /a/ maps onto the non-peripheral [ɐ]; (iii) each of /ɛ, e/ also map onto the central [ə]. Although /ɛ, e/ are not corner vowels, they are more peripheral than schwa. These two mappings reduces the contrast between vowels in unstressed syllables – rather than being pushed apart in the articulatory space, the vowels centralize, moving towards each other in the space. I will therefore analyze the vowel reduction in Portuguese as a prominence reduction process.

Aside from these reduction processes, European Portuguese also has variable vowel deletion in unstressed syllables. Vowel deletion is usually not treated in as much detail as vowel reduction in the literature on Portuguese grammar. Mateus and d'Andrade, for instance, devote only one paragraph to vowel deletion (2000:18), while vowel

---

[10] Crosswhite (2000b:2) makes a similar claim. Schwa is usually very short and low in intensity in comparison to other vowels. Also, although schwa is usually classified as a mid vowel, Pettersson and Wood (1987) have found that, at least in Bulgarian, schwa is pronounced with a very close jaw position – similar to that seen in high vowels such as /i, u/.

[11] I am classifying all of the mid, non-central vowels together. This is because there is no evidence that they are treated differently with regard to the vowel reduction or vowel deletion. Since these vowels are all contiguous on the sonority hierarchy, this is simply scale conflation or encapsulating (de Lacy, 2003a, 2003b, Prince and Smolensky, 1993).

reduction receives several sections. As far as the specific pattern of variation is concerned, Mateus and d'Andrade make only three remarks: (i) It is only schwa[12] and unstressed [u] that are subject to deletion; (ii) schwa is more prone to delete than [u]; and (iii) deletion is mostly limited to word final position.

Silva (1997, 1998) represents the first detailed study of the process of vowel deletion. For Faialense Portuguese he shows that: (i) Although schwa and [u] delete most frequently, other vowels in unstressed position can also delete; (ii) word-final vowels are more prone to deletion than non-word final vowels; (iii) a vowel is more likely to delete if the following syllable is unstressed than if the following syllable is stressed. In general then, Silva has shown that the deletion process is more wide spread than what has traditionally been assumed, and that the rate of deletion is at least partially determined by grammatical factors. I will report Silva's findings in more detail in §3 where I present an OT account of this process.

2. Vowel reduction in Faialense Portuguese

In this section I provide an account for vowel reduction in Faialense Portuguese within the theory of vowel reduction developed by Crosswhite (2000a, 2000b, 2001). This section is structured as follows: I first discuss the constraints that are necessary to account for the reduction process (§2.1). After that, I show how these constraints can be used to account for vowel reduction in Faialense Portuguese (§2.2 to §2.7).

---

[12] Of course, Mateus and d'Andrade do not recognize schwa as part of the vowel inventory of Portuguese. In the discussion here I have replaced all of their references to [ɨ] with [ə]. See footnote 5 above.

2.1 The constraints

Crosswhite offers an explanation of prominence neutralizing vowel reduction by appealing to the concept of harmonic alignment as formulated by Prince and Smolensky (1993). The basic idea behind harmonic alignment is that different types of prominence should be aligned with each other. Assuming two prominence scales, harmonic alignment requires prominent elements from one scale to co-occur with prominent elements from the other scale, and similarly for non-prominent elements on the two scales. Prominence neutralizing vowel reduction aims to have non-prominent (low sonority) vowels align with prosodically weak positions (unstressed syllables). This is therefore exactly the situation in which harmonic alignment can be called upon. In (9) I list the two prominence scales involved in Faialense Portuguese vowel reduction, and show their harmonic alignment.

(9) **Harmonic alignment of syllable strength and sonority**

Syllabic prominence scale:	σ́ > σ̆
Sonority scale for vowels:	a > ɐ > {o, ɔ, e, ɛ} > u > i > ə
Harmonic alignment for σ̆:	σ̆/ə ™ σ̆/i ™ σ̆/u ™ σ̆/{o, ɔ, e, ɛ} ™ σ̆/ɐ ™ σ̆/a
Harmonic alignment for σ́:	σ́/a ™ σ́/ɐ ™ σ́/{o, ɔ, e, ɛ} ™ σ́/i ™ σ́/u ™ σ́/ə

Since vowel reduction occurs in *unstressed* syllables, the harmonic alignment of these two scales in terms of σ́ is not relevant here, and it will not be discussed any further.[13] These harmonically aligned scales can now be converted into constraints. The

[13] Portuguese does not have the converse of vowel reduction in unstressed syllables, that is, vowel augmentation in stressed syllables – i.e. it is not the case that, for instance, an underlying /u/ that is parsed into a stressed syllable is replaced by a vowel of higher sonority such as [o]. Therefore, the

constraints will be in a fixed ranking relationship determined by the harmonic alignment. The members of the harmonically aligned ŏ-scale are ordered from the most to the least well-formed. The ranking between the constraints is therefore the opposite of the ordering between the elements on the harmonically aligned scale, so that the least well-formed member on the scale will violate the highest ranked constraint. In (10) I list the constraints that can be derived from the harmonically aligned ŏ-scale. Rather than listing the mid vowels individually, I refer to {o, ɔ, e, ɛ} together as "mid". Note that this group does not include schwa even though schwa is also a mid vowel. These vowels are therefore not really the mid vowels, but rather a group of vowels that occupy the same slot on the sonority scale.

(10) **Prominence alignment constraints**

||*ŏ/a ○ *ŏ/ɐ ○ *ŏ/mid ○ *ŏ/u ○ *ŏ/i||

An /a/ vowel in an unstressed syllable will violate the highest ranked constraint *ŏ/a, while an /i/ in an unstressed syllable will violate the lowest ranked constraint *ŏ/i. There is no constraint against parsing schwa into an unstressed syllable. This is in accordance with Gouskova (2003) who shows that there can be no constraints against the least marked member on a harmonically aligned scale. It is the interaction of these markedness constraints with faithfulness constraints on featural identity that determines which vowels are reduced and to what they reduce.

---

constraints that can be formed from the harmonic alignment on σ́ have to be ranked very low in the constraint hierarchy of Portuguese.

There are two ways in which featural identity constraints can be stated, namely as directional IDENT[±F] faithfulness constraints (Pater, 1999), or as the more traditional non-directional IDENT[F] constraints (McCarthy and Prince, 1994, 1995). The directional IDENT[±F] constraints differ from the non-directional IDENT[F] constraints in that they are able to distinguish between the mappings /+F/ → [-F], and /-F/ → [+F]. The definitions of these two versions of IDENT constraints are given in (11) and are based on Pater (1999) and McCarthy and Prince (1994, 1995).[14]

(11) **a. Non-directional**

IDENT[F]

If *x* is an output correspondent of an input segment *y*, then *x* must agree with *y* in its specification for the feature [F].

b. Directional

IDENT[+F]

If *x* is an output correspondent of an input segment *y* and *y* is specified as [+F], then *x* must also be specified as [+F].

IDENT[-F]

If *x* is an output correspondent of an input segment *y* and *y* is specified as [-F], then *x* must also be specified as [-F].

[14] There is actually a third way in which featural identity can be enforced, and that is through MAX[F]/DEP[F] constraints (Lombardi, 1998, 2001). These constraints are also violated by ordinary segmental deletion and epenthesis. Since Portuguese phonology needs to distinguish between featural change (reduction) and deletion, I will not use these MAX[F]/DEP[F] constraints here.

To illustrate the difference between these two approaches to featural faithfulness constraints, consider tableau (12) in which I use IDENT constraints for the feature [high].

(12) **Comparison between different kinds of featural faithfulness**

	Non-Directional	Directional	
	IDENT[high]	IDENT[+high]	IDENT[-high]
/e/ → [i]	*		*
/i/ → [e]	*	*	

The non-directional IDENT constraint cannot distinguish between raising and lowering, while the directional constraints can. The typology predicted by the non-directional constraints therefore forms a subset of the typology predicted by the directional constraints. In a grammar where the directional constraints for a feature [F] are ranked contiguously (no constraints intervening between them), the same output will be selected even if the directional constraints were replaced by a single non-directional constraint. It is only when the two directional constraints are separated by other constraints that the predictions of the two approaches diverge. A non-directional constraint can therefore be seen as shorthand for two directional constraints that are contiguously ranked.

Pater (1999) has shown with examples from nasalization and de-nasalization in Austronesian that directional constraints are necessary. I will therefore use directional constraints here. However, in the vocalic phonology of Portuguese, it is only for the feature [high] that it is necessary to rank constraints in between the two directional IDENT constraints. For all the other features, the directional constraints can be ranked contiguously. In order to simplify the exposition below, I will use non-directional constraints for all features except for [high]. For each of the other features, the non-

directional constraints can be replaced with directional constraints ranked contiguously. I list the featural faithfulness constraints that I will use in (13).

(13) **Featural faithfulness constraints active in Portuguese vowel reduction**

IDENT[front]

IDENT[back]

IDENT[low]

IDENT[ATR]

IDENT[+high], IDENT[-high]

In §1.2 above I have already discussed the data that needs to be explained. However, I am repeating the essential aspects of the vowel reduction process again in (14) to facilitate the discussion in the following sections.

(14) **What needs to be explained**

a. Low vowels: /a, ɐ/ → [ɐ̆]

b. Back vowels: /o, ɔ, u/ → [ŭ]

c. Schwa:[15] /ə/ → [ə̆]

d. High front vowel: /i/ → [ĭ]

e. Mid front vowels: /e, ɛ/ → [ə̆]

[15] Schwa does not have phonemic status in Portuguese – the only surface schwa's are the result of vowel reduction. However, under richness of the base (Prince and Smolensky, 1993, Smolensky, 1996) we also have to consider how a schwa input will be treated. Since a schwa is the preferred unstressed syllable, there would not be any pressure on an underlying schwa to change if it is parsed into an unstressed syllable. I am therefore making the assumption that schwa will map faithfully onto the surface.

Before delving into the details of the vowel reduction process, I explain in §2.1.1 a basic assumption that I make about the ranking between markedness and faithfulness constraints. In the sections §2.2 to §2.7 I then discuss each of the classes of vowels mentioned in (14). I am making two simplifying assumptions in the discussion: (i) I am ignoring the deletion candidate; and (ii) I am not taking into account the possibility of variation. I will deal in detail with both of these issues in §3 and §4 below.

2.1.1 Ranking conservatism

Throughout the discussion here I follow the principle of "ranking conservatism" (Itô and Mester, 1999, 2003, Tesar and Smolensky, 1998, 2000) . This principle is based on the assumption that the original state of the grammar has the ranking ||Markedness o Faithfulness|| (Smolensky, 1996). I will therefore assume this ranking between any markedness constraint and faithfulness constraint unless if there is positive evidence to the contrary. This is not a necessary assumption – the vowel reduction and deletion process in Faialense Portuguese can be explained without this assumption. However, there are two reasons for making this assumption: First, it is in agreement with standard assumptions about grammar in the OT literature. Secondly, it results in a "neater" looking final grammar – a grammar that more closely approaches a total ranking of the constraints. One reason for this assumption is therefore aesthetical.

2.2 Low vowels

Both low vowels map onto the vowel [ɐ̆] in unstressed syllables, i.e. /a, ɐ/ → [ɐ̆]. Consider first the mapping /a/ → [ɐ̆]. This observed unfaithful mapping violates IDENT[ATR], while the faithful mapping violates *ɔ̆/a. In order for the unfaithful

candidate to be preferred over the faithful candidate, we therefore need the ranking ||*ŏ/a o IDENT[ATR]||.

We still need to explain why /a/ maps onto [ɐ̆] rather than onto some other unfaithful candidate. The actually observed output [ɐ̆] violates the markedness constraint *ŏ/ɐ. If /a/ mapped onto any of the non-low vowels, it would have violate a lower ranked markedness constraint. This implies the mapping of /a/ onto any non-low vowel should violate a faithfulness constraint ranked higher than *ŏ/ɐ. The only constraint violated by the mapping /a/ → [ə̆] is IDENT[low], so that we need at least the ranking ||IDENT[low] o *ŏ/ɐ|| in order to prevent /a/ from mapping onto [ə̆]. It so happens that all other mappings of /a/ to non-low vowels will also violate IDENT[low]. This ranking alone is therefore sufficient to rule out all unfaithful mappings of /a/ except for the actually observed /a/ → [ɐ̆]. The full ranking required to explain the mapping /a/ → [ɐ̆] is therefore ||{*ŏ/a, IDENT[low]} o {*ŏ/ɐ, IDENT[ATR]}||. This is shown in the tableau in (15).

(15) **/a/ → [ɐ̆]**[16]

/a/	*ŏ/a	ID[low]	*ŏ/ɐ	ID[ATR]
ă	*!			
ɐ̆			*	*
All other vowels		*!		(*)[17]

[16] The rankings ||*ŏ/a o IDENT[low]|| and ||*ŏ/ɐ o IDENT[ATR]|| are not necessary to explain the mapping /a/ → [ɐ̆]. I assume these rankings based on the principle of "ranking conservatism". See this discussion in §2.1.1 above.

[17] IDENT[ATR] is violated by all [+ATR] vowels and obeyed by all [-ATR] vowels. However, since all of these vowels also violate high ranking IDENT[low], their performance on IDENT[ATR] is not relevant. Throughout the discussion I will use a parenthesized asterisk (*) to indicate a constraint that is violated by some but not all candidates represented by a row in a tableau.

The faithful candidate [ă] fatally violates *ŏ/a. All unfaithful candidates except for [ĕ] fatally violate IDENT[low]. Consequently, [ĕ] is the observed output for an /a/-input.

Now consider /ɐ/ → [ĕ]. This mapping violates only the constraint *ŏ/ɐ. If /ɐ/ were to map onto its [-ATR] counterpart [ă] it will violate a higher ranked markedness constraint *ŏ/a. All mappings of /ɐ/ onto a non-low vowel will violate the constraint IDENT[low], which, as we have already established, outranks *ŏ/ɐ. No additional rankings are necessary to explain why /ɐ/ does not reduce. This is shown in tableau (16).

(16) **/ɐ/ → [ĕ]**

/ɐ/	*ŏ/a	ID[low]	*ŏ/ɐ	ID[ATR]
ă	*!			*
ĕ			*	
All other vowels		*!		(*)

2.3 Back vowels

All back vowels map onto the high back vowel in unstressed syllables, i.e. /o, ɔ, u/ → [ŭ]. Consider first the mapping of the mid back vowels, /o, ɔ/ → [ŭ]. Had these vowels been mapped faithfully onto the surface, they would have violated the markedness constraint *ŏ/mid. The actually observed mappings violate *ŏ/u and either only IDENT[-high] (/o/ → [ŭ]) or both IDENT[-high] and IDENT[ATR] (ɔ/ → [ŭ]). This means that *ŏ/mid has to outrank *ŏ/u, IDENT[-high] and IDENT[ATR], i.e. ||*ŏ/mid o {*ŏ/u, IDENT[-high], IDENT[ATR]}||.[18] This explains why these inputs do not map faithfully onto themselves.

[18] The ranking ||*ŏ/mid o *ŏ/u|| was motivated in (10). Only the other two rankings are therefore new.

However, this alone still does not explain why /o, ɔ/ does not map onto [ĭ] or [ə̆]. Both of these vowels are lower in sonority than [ŭ], and therefore violate lower ranked markedness constraints from the hierarchy in (10) than does [ŭ]. The mapping /ɔ/ → [ə̆] violates only the faithfulness constraint IDENT[back]. This means that the ranking ||IDENT[back] ○ *ŏ/u|| is required to block this mapping. The mappings /o/ → [ə̆], and /ɔ, o/ → [ĭ] all also violate IDENT[back]. The ranking ||IDENT[back] ○ *ŏ/u|| is therefore sufficient to block all of these mappings.[19]

The following ranking is therefore necessary to explain the mappings /o, ɔ/ → [ŭ]: ||{*ŏ/mid, IDENT[back]} ○ *ŏ/u ○ {IDENT[-high], IDENT[ATR]}||. This is illustrated in the tableau in (17).[20]

(17) **/ɔ/ → [ŭ]**[21]

/ɔ/	*ŏ/mid	ID[ba]	*ŏ/u	ID[-hi]	ID[ATR]
ɔ̆	*!				
ŏ	*!				*
ŭ			*	*	*
All other vowels		*!		(*)	(*)

[19] /o/ → [ə̆] also violate IDENT[ATR]. This mapping can therefore also be blocked by ranking IDENT[ATR] over *ŏ/u. Similarly /ɔ/ → [ĭ] also violates IDENT[ATR] and IDENT[front]. This mapping can therefore also be blocked by ranking either IDENT[ATR] or IDENT[front] over *ŏ/u. Lastly, the mapping /o/ → [ĭ] also violates IDENT[front], and can therefore also be blocked by ranking IDENT[front] over *ŏ/u. The ranking argued for in the text, ||IDENT[back] ○ *ŏ/u||, is therefore sufficient but not necessary. However, following the principle of ranking conservatism (see §2.1.1) I am opting for the ranking ||IDENT[back] ○ *ŏ/u||, since this ranking eliminates the need for any of IDENT[ATR] or IDENT[front] to be ranked above a markedness constraint. This ranking is the most conservative ranking that can explain the data.

[20] I do not include as candidates any vowels that are lower in sonority than /o, ɔ/ – i.e. I do not include the candidates [ă, ĕ]. Since they are higher in sonority than /o, ɔ/, they violate higher ranked markedness constraints than the faithful candidate – see (10) above. They can therefore not be selected as output candidates.

[21] The rankings ||*ŏ/mid ○ IDENT[back]|| and ||*ŏ/u ○ {IDENT[-high], IDENT[ATR]}|| are based on the principle of ranking conservatism – see §2.1.1 above.

The observed output [ŭ] violates *ŏ/u, IDENT[-high] and IDENT[ATR]. The faithful candidate [ɔ̆] violates only *ŏ/mid. However, because *ŏ/mid dominates all of the constraints violated by the observed candidate [ŭ], the faithful candidate is eliminated. All unfaithful candidates except for [ŭ] violate IDENT[back]. This is a fatal violation because IDENT[back] outranks *ŏ/u. The result is that [ŭ] is the optimal candidate. In this tableau I used /ɔ/ as input. However, the same point can be made also with an /o/-input. The only difference will be that [ɔ̆] rather than [ŏ] will violate IDENT[ATR].

Now consider the high back vowel that maps faithfully onto itself, /u/ → [ŭ]. Mapping onto any of [ă, ɐ̆, ŏ, ɔ̆, ĕ, ɛ̆] violates markedness constraints ranked higher than the markedness constraint *ŏ/u, which is violated by the faithful candidate [ŭ] – see (10) above. There are therefore only two candidates to worry about, namely [ə̆, ĭ]. Both of these candidates violate the faithfulness constraint IDENT[back] which has already been established to outrank *ŏ/u. No additional rankings are therefore necessary to explain why /u/ does not reduce. This is illustrated in (18).

(18) **/u/ → [ŭ]**

/u/	*ŏ/mid	ID[ba]	*ŏ/u	ID[-hi]	ID[ATR]
ɔ̆	*!				*
ŏ	*!				
ŭ			*		
All other vowels		*!		(*)	(*)

The faithful candidate violates only *ŏ/u. All other candidates violate either *ŏ/mid or IDENT[back], both of which outrank *ŏ/u. The faithful candidate is therefore selected as output.

2.4 Schwa

Since schwa is the preferred vowel in an unstressed syllable, there is no pressure on an underlying schwa to change into any other vowel if it is parsed into an unstressed syllable. In fact, there is not even a markedness constraint in the markedness hierarchy in (10) that penalizes an unstressed schwa. The fully faithful mapping /ə/ → [ə̆] therefore violates none of the markedness constraints under consideration here. This means that no unfaithful candidate can improve in markedness on the fully faithful candidate. The principles of harmonic ascent (Moreton, 1999) and harmonic bounding (Samek-Lodovici and Prince, 1999) therefore assure that /ə/ will map faithfully onto [ə̆]. No tableau is given to illustrate this. The faithful candidate violates neither a markedness nor a faithfulness constraint. Each of the other candidates violates one markedness constraint and at least one faithfulness constraint. It is therefore clear that the faithful candidate will be optimal.

2.5 The high front vowel

The high front vowel /i/ maps faithfully onto itself, i.e. /i/ → [ĭ]. This faithful mapping violates the markedness constraint *ŏ/i. With the exception of [ə̆] all other possible candidates violate a markedness constraint that outranks *ŏ/i and are non-optimal. The mapping /i/ → [ə̆] is therefore the only mapping that needs to be ruled out by a faithfulness constraint. This mapping violates three faithfulness constraints, namely IDENT[+high], IDENT[front] and IDENT[ATR]. As long as one of these constraints ranks higher than *ŏ/i, the reduction mapping will be blocked. I will assume here that it is the

constraint IDENT[+high] that is acting as the blocking constraint.[22] All other unfaithful mappings will violate markedness constraints that are ranked higher than *ŏ/i. They can therefore never beat the faithful candidate. The only new ranking required is ||IDENT[+high] o *ŏ/i||. This is shown in the tableau in (19). I do not include in this tableau any of the candidates that violate a markedness constraint from (10) ranked higher than *ŏ/i.

(19) **/i/ → [ĭ]**

/i/	ID[+hi]	*ŏ/i
ĭ		*
ə̆	*!	

2.6 The mid front vowels

The mid front vowels /e, ɛ/ reduce to [ə̆] in unstressed syllables. The faithful candidates for these two vowels violate the markedness constraint *ŏ/mid. It is therefore necessary that all faithfulness constraints violated by the actual mapping /e, ɛ/ → [ə̆] be ranked lower than *ŏ/mid. The mapping /e/ → [ə̆] violates IDENT[ATR] and IDENT[front], while the mapping /ɛ/ → [ə̆] violates only IDENT[front]. The following ranking is therefore minimally necessary: ||*ŏ/mid o {IDENT[front], IDENT[ATR]}||.

[22] This is not a necessary assumption. Any one or combination of these constraints could act as blocker. Under the conservative assumption that faithfulness constraints will be ranked as low as possible (§2.1.1), I will assume that only one of them actually ranks above *ŏ/i. Both IDENT[front] and IDENT[ATR] are violated elsewhere in Portuguese. IDENT[front] is violated in the reduction of the mid front vowels /e, ɛ/ to [ə̆] (see §2.6 below), and IDENT[ATR] is violated by several mappings, for instance /ɔ/ → [ŭ] (see §2.3 above). Since these constraints are at least sometimes violated while IDENT[+high] is never violated, the most conservative option is to let IDENT[+high] be the blocker.

The vowels [ă, ɐ̆, ŏ, ɔ̆] all violate a markedness constraint from (10) that is ranked as least as high as the markedness constraint *ŏ/mid violated by the faithful [ĕ, ɛ̆]. These candidates can therefore not be optimal. The vowels [ĭ, ŭ] violate *ŏ/i and *ŏ/u respectively. Both of these are ranked lower than *ŏ/mid – see (10). We can prevent /e, ɛ/ from mapping onto [ĭ, ŭ] by ranking the faithfulness constraint violated by /e, ɛ/ → [ə̆] lower than *ŏ/i and *ŏ/u. This ranking is also in accordance with the principle of raking conservatism – see §2.1.1.

The rankings necessary to explain the mapping /e, ɛ/ → [ə̆] is therefore ||*ŏ/mid ○ *ŏ/u ○*ŏ/i ○ {IDENT[front], IDENT[ATR]}||. This is shown in the tableau in (20). This tableau does not include any of the candidates that are more marked than the faithful candidate.

(20) **/e/ → [ə̆]**[23]

/e/	*ŏ/mid	*ŏ/u	*ŏ/i	ID[ATR]	ID[fr]
ə̆				*	*
ĕ	*!				
ŭ		*!			*
ĭ			*!		

The observed output [ə̆] violates only the faithfulness constraints IDENT[ATR] and IDENT[front]. The faithful candidate [ĕ] violates *ŏ/mid. Because *ŏ/mid ranks higher than IDENT[ATR] and IDENT[front], this violation is fatal. The vowels [ĭ, ŭ] violate *ŏ/i

[23] When I discuss deletion later in this chapter, I will show that the situation with the /e, ɛ/ is more complicated than what is presented here. In particular, I will show that it is not the *ŏ/i that prevents /e, ɛ/ from mapping onto [ĭ], but rather the local conjunction of IDENT[-high] and *ŏ/i. However, the evidence for this comes from the variable deletion of /e, ɛ/, and since the data on deletion have not yet been presented, I cannot yet motivate this complication. See the discussion in §3.3.1 below.

and *ŏ/u respectively. Since these two constraints also outrank IDENT[ATR] and IDENT[front], [ĭ, ŭ] are ruled out as candidates. The illustration was here given in terms of an /e/-input. However, the same can be shown with an /ɛ/-input. The only difference will be that the observed output [ə̆] will not violate IDENT[ATR], and [ɛ̆] rather than [ĕ] will violate *ŏ/mid.

2.7 Summary

The rankings necessary to explain vowel reduction is Faialense Portuguese are summarized in the table in (21). In this table I indicate the ranking in the first column, and the motivation for the ranking in the second column. The third column indicates where in the preceding discussion that particular ranking is motivated. After the table, I give a graphic representation of these rankings.

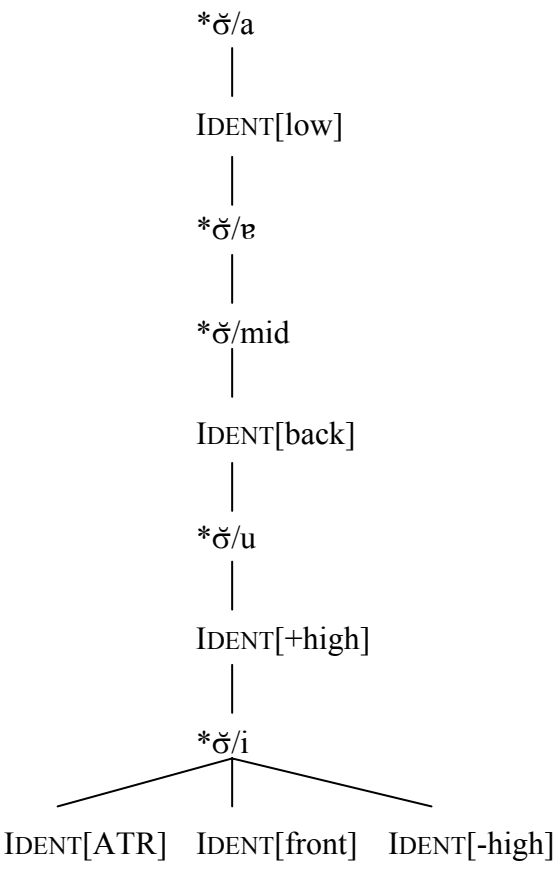
(21) **Summary of rankings necessary for vowel reduction**

Ranking	**Motivation**	**Where?**
*ŏ/a o*ŏ/ɐ o*ŏ/mid o *ŏ/u o*ŏ/i	Universal	§2.1 (10)
*ŏ/a o IDENT[ATR]	/a/ reduces to [ɐ̆]	§2.2 (15)
IDENT[low] o *ŏ/ɐ	/a, ɐ/ do not reduce to non-low vowels	§2.2 (15)
*ŏ/mid o IDENT[-high], IDENT[ATR]	/o, ɔ/ reduce to [ŭ]	§2.3 (17)
IDENT[back] o *ŏ/u	/o, ɔ, u/ do not reduce to [ĭ] or [ə̆]	§2.3 (17)
IDENT[+high] o *ŏ/i	/i/ does not reduce to [ə̆]	§2.5 (19)
*ŏ/mid o IDENT[front], IDENT[ATR]	/e, ɛ/ reduce to [ə̆]	§2.6 (20)
*ŏ/u, *ŏ/i o IDENT[front], IDENT[ATR]	/e, ɛ/ do not reduce to [ĭ] or [ŭ]	§2.6 (20)

((21) continued)

Ranking	**Motivation**	**Where?**
*ŏ/a o IDENT[low]	Ranking conservatism	§2.1.1
*ŏ/mid o IDENT[back]		
*ŏ/u o IDENT[+high]		
*ŏ/i o IDENT[-high], IDENT[front], IDENT[ATR]		

(22) **Graphic representation of the rankings for Faialense Portuguese vowel reduction**



3. The interaction of vowel reduction and deletion in Faialense Portuguese

Aside from vowel reduction, European Portuguese also applies a more severe form of vowel lenition in unstressed syllables, namely vowel deletion. However, unlike reduction,

deletion is a variable process. A vowel in an unstressed syllable is sometimes pronounced and sometimes deleted. This means that a single word often has more than one possible pronunciation. There are two possibilities: (i) If the unstressed vowel is a vowel that is subject to reduction, then the variation will be between reduction and deletion. (ii) If the unstressed vowel is a vowel that resists reduction, then the variation will be between faithfully preserving the underlying vowel and deletion.

(23) **Variation between the faithful, reduction and deletion candidates**

Reduction ~ Deletion:	/peludo/	→	[pəlúdŭ]~[p_lúdŭ]	*peludo*	"hairy"
Faithful ~ Deletion:	/piloto/	→	[pĭlótŭ]~[p_lótŭ]	*piloto*	"pilot"

In this and the next section I will offer an OT account for this variable pattern within the rank-ordering model of EVAL. This section is structured as follows. In §3.1 I present that data on vowel deletion in Faialense Portuguese. In §3.2 I develop an OT account for the differential realization of each of the individual vowels. This is what I refer to as intra-contextual variation – any given input can be pronounced in more than one manner (see Chapter 1 §2.2.1). In the next section (§4), I account for gross patterns of the variable process that hold true across different vowels – in general vowels are more or less prone deletion based on the context in which they appear. This is what I refer to inter-contextual variation (see Chapter 1 §2.2.2).

3.1 The data

Mention of variable vowel deletion is found throughout the literature on European Portuguese (Crosswhite, 2001:104, Mateus and d'Andrade, 2000:18, Oliviera, 1993:9, Parkinson, 1988:142). However, the discussion of vowel deletion typically amounts to no

more than an acknowledgement that the process exists. The only exceptions to this generalization are two papers by Silva. He reports on unstressed vowel deletion in two dialects of Azorean Portuguese, namely that spoken on the island of Faial (Silva, 1997) and that spoken on the island of São Miguel (Silva, 1998). In these two papers Silva shows that the pattern of deletion is not random, but that it is at least partially determined by grammatical factors. He offers an account of this process within the variable rule framework of Labov (Cedergren and Sankoff, 1974, Kay and McDaniel, 1979, Labov, 1972). In particular he employs the VARBUL software package (Sankoff and Rand, 1988) to determine which grammatical factors contribute significantly towards determining the observed pattern of deletion. I will not adopt Silva's variable rule analysis of vowel deletion, but I will use the data on the deletion process as he presents it. Although the deletion patterns in Faialense and São Miguel Portuguese are very similar, they are not identical. In the rest of this discussion I will focus only on Faialense Portuguese. I choose Faialense Portuguese over São Miguel Portuguese since Faialense Portuguese is very similar to standard (Lisbon) European Portuguese while São Miguel Portuguese differs much more from the European standard.[24]

---

[24] See footnote 3 about the just how similar Faialense Portuguese is to standard European Portuguese. In that footnote I explain that Faialense Portuguese agrees with standard European Portuguese in all respects relevant to vowel reduction and deletion.

São Miguel Portuguese, on the other hand, is more different from standard European Portuguese. For instance, Silva (1997:30, note 3) claims that São Miguel Portuguese has a different vowel inventory than standard European Portuguese. Also see Rogers (1949:48) who claims that the Portuguese of the eastern Azores is quite different in pronunciation from Lisbon Portuguese. São Miguel is an eastern Azorean island.

Silva collected his data on Faialense Portuguese from a 42 year old female native speaker of Faialense Portuguese.[25] He recorded a conversation between the subject and her mother, and selected a continuous block of 22 minutes for study. He transcribed this block, and then analyzed the realization of vowels in unstressed syllables.[26] This resulted in 861 unstressed syllables or potential sites for vowel deletion.[27] Silva reported the data without reference to the underlying form of the vowels. For instance, the data that he reports on deletion of [ŭ] includes [ŭ] that corresponds to underlying /o, ɔ, u/ (see §2.3 above on this). This is a little unfortunate, as it is at least conceivable that different underlying forms might be subject to different rates of deletion. However, because of the way in which Silva reports his data, I will make the assumption that the deletion rates reported for [ŭ] apply equally to [ŭ] derived from each of /o, ɔ, u/.

Silva coded his data according to several linguistic factors. These data were then submitted to the VARBUL program (Sankoff and Rand, 1988). From among all the factors that Silva considered, only three were selected by VARBUL as contributing significantly

---

[25] The conclusions drawn here about Faialense Portuguese are therefore somewhat tentative, being based on the speech of a single speaker – this is also acknowledged by Silva (1997:307, endnote 7). However, under the assumption that the data presented by Silva reflect at least the grammar of this individual, it is necessary that our theory be capable of accounting for the data.

[26] Since vowel deletion results in loss of a syllable, the term "unstressed syllable" should not be given a strictly phonetic meaning here. For each word, Silva determined which syllables would have been unstressed had all the vowels been pronounced. He counted as unstressed syllables therefore also those syllables that were destroyed by the deletion of the vowels that would have formed their nuclei.

[27] Originally Silva assumed that there were 884 potential sites of vowel deletion. However, upon closer scrutiny of the data, he found that the vowel in the last syllable of the third person pronouns *elle*/*elles* and the vowel in the first syllable of the preposition *para* were never realized. Based on this he assumed that the underlying forms of the third person pronouns do not contain a vowel in the second "syllable", i.e. /el/ and /els/. And similarly he assumed that the underlying form of *para* does not actually contain a vowel in the first "syllable", i.e. /pra/. With regard to *para* there is additional evidence from Brazilian Portuguese that the underlying form is /pra/ – unlike European Portuguese, Brazilian Portuguese is not characterized by deletion of unstressed vowels, and even so *para* in Brazilian Portuguese is pronounced as [pra] (Thomas, 1974:15). The 23 occurrences of these three lexical items were excluded from the final analysis (Silva, 1997:303-304).

toward determining how likely vowel deletion is to apply. These three factors are: (i) vowel quality (i.e. some vowels are more likely to delete than others), (ii) position in prosodic word (word final vowels are more likely to delete), and (iii) stress of following syllable (vowels are more likely to delete if followed by an unstressed syllable). Table (24) summarizes the deletion pattern and is based on *Table 2* from Silva (1997:305).[28]

(24) **Vowel deletion patterns in Faialense Portuguese**

		Following syllable				
		Stressed		**Unstressed**		
		ω-final	**Elsewhere**	**ω-final**	**Elsewhere**	**Total**
[ə̆]	**Deleted**	12	23	20	13	68
	n	19	86	26	32	163
	% deleted	63%	27%	77%	41%	42%
[ŭ]	**Deleted**	23	7	25	9	64
	n	35	65	35	47	182
	% deleted	66%	11%	71%	19%	35%
[ĭ]	**Deleted**	—	5	—	—	5
	n	—[29]	75	—	—	75
	% deleted	—	7%	—	—	7%
[ĕ]	**Deleted**	1	3	1	2	7
	n	75	147	58	51	331
	% deleted	1%	2%	2%	4%	2%

[28] Silva includes nasal vowels, but as explained earlier my focus is only on the oral vowels and therefore I will not include his data on the nasal vowels (cf. §1.1). He also includes another prosodic context, namely pre-pausal. However, this factor was never selected by the VARBUL program as a significant factor (Silva, 1997:304). I am therefore also not including the data on this prosodic position.

[29] Silva lists 10 occurrences of [ĭ] in this cell. However, in personal communication to me Silva explained these 10 [ĭ]'s were all occurrences of the conjunction *e* 'and'. In his coding of the data he treated this word as a separate prosodic word. This is probably not correct. Function words are incorporated into the prosodic word headed by the following lexical word (see below). These 10 [ĭ]'s should rather be counted as [ĭ]'s occurring non-final in a prosodic word followed by a stressed vowel. This is indeed how I represent them in this table – i.e. the 75 [ĭ]'s in the next cell include 10 [ĭ]'s that in Silva's original table were in this cell. See also the discussion footnote 8 about post-tonic /i/.

Vowel quality. Not all vowels are equally likely to delete. The vowels are ordered as follows according to how likely they are to delete: [ə̆] > [ŭ] > [ĭ] > [ɐ̆].[30] The total deletion rates of [ə̆] and [ŭ] do not differ significantly, although it tends strongly towards significance ($\chi^2(1) = 3.21$, $p = .07$).[31] The deletion rates of [ŭ] and [ĭ] do differ significantly ($\chi^2(1) > 10^{307}$), as do deletion rates of [ĭ] and [ɐ̆] ($\chi^2(1) = 11.02$, $p = .0009$). The deletion rate of [ɐ̆] does differ significantly from zero ($p = .0008$).[32] However, since it is so close to zero, I will treat [ɐ̆] as if it is different from all the other vowels. All the other vowels I will treat as if variable deletion is indeed attested for them. However, [ɐ̆] will be treated as if it never deletes.

Position in prosodic word. Vowels that are final in a prosodic word are consistently more likely to delete than vowels that occur elsewhere in a prosodic word. The difference in deletion rates between these two contexts is significant. Of the 115 unstressed vowels that occurred in prosodic word final position 80 or 70% were deleted. However, of the 305 unstressed vowels that occurred elsewhere in the phonological word,

---

30 This pattern agrees with that found by Cedergren and Simoneau (1985) for Montréal French in which non-low vowels are more likely to delete than low vowels.

31 These statistics were calculated as follows: As observed frequencies I took the number of deletions and retentions of the vowel with the lower deletion rate. In the comparison between [ə̆] and [ŭ], I therefore took the number of deletions and retentions of [ŭ] as observed. The expected frequencies were then calculated by assuming that the vowel with the lower deletion rate actually had the same deletion rate as the vowel with the higher deletion rate – i.e. I assumed [ŭ] had a deletion rate of 42% rather than its actual rate of 35%. The frequencies used for the comparison between [ə̆] and [ŭ] were therefore the following:

	Observed	**Expected**
Deletion	64	(0.42)(182) = 76
Retention	118	(0.58)(182) = 106

32 This is the binomial probability of having 0 successes out of 331 trials if the probability of success on every trial is really 0.02.

only 57 or 19% were deleted ($\chi^2(1) > 10^{307}$).[33] Here it is relevant to know how Silva determined the boundaries of prosodic words. Following Selkirk (1986, 1987), he assumes that prosodic words are defined in part by reference to syntactic structure (Silva, 1997:297). In particular, he claims that every lexical word in a syntagmatic sequence corresponds to a prosodic word, and that it is the right edge of the lexical word that is used to delimit the boundary between prosodic words. This implies that functional words are incorporated into the prosodic word headed by a following lexical word. In a sentence such as *As mulheres de Coimbra cantavam um fado* 'The women from Coimbra sang a fado', there are then four prosodic words: [*As mulheres*]$_\omega$ [*de Coimbra*]$_\omega$ [*cantavam*]$_\omega$ [*um fado*]$_\omega$. With the expression "final in prosodic word", Silva indicates a vowel that is final in a prosodic word, not the final vowel in a prosodic word. The last vowel in *Coimbra* was coded as "final in prosodic word", but the last vowel of *mulheres* was coded as "elsewhere in prosodic word".

Stressed or unstressed following syllable. An unstressed vowel is more likely to delete when followed by an unstressed syllable than when followed by a stressed syllable. Of the 140 unstressed vowels followed by an unstressed syllable 67 (48%) deleted. Of the 280 unstressed vowels followed by a stressed syllable 70 (25%) deleted ($\chi^2(1) > 10^{307}$).[34]

There are two aspects to the variation that needs to be accounted for. (i) *Intra-contextual variation*. The relationship between deletion and reduction for individual vowels needs to be explained – e.g. why for a some vowel deletion or reduction is more

---

[33] In these calculations I include only the vowels that are subject to deletion (i.e. only potential sites for deletion). The low vowels are therefore not included in these counts.

[34] See the previous footnote.

or less frequent. This is the focus of the current section. (ii) *Inter-contextual variation*. But the differences between contexts across vowels also need to be accounted for – why is deletion more frequently for some vowels than others, why does deletion occur more in pre-unstressed than pre-stressed position, and why do the deletion rates differ for vowels that occur final in prosodic words and for vowels that occur elsewhere in prosodic words. This will be discussed in the section §4.

Although this section is dedicated to variation in the realization of individual vowels, the position of the vowel in the prosodic word will also be discussed. This is necessary because the relative frequency of deletion and retention differs for some vowels depending on where they occur in the prosodic word. For instance, final in a prosodic word [ŭ] is preferentially deleted (69%), but elsewhere in the prosodic word deletion is dispreferred (14%). Throughout the discussion below I will use the symbol ∅ to stand for a candidate in which the unstressed vowel has been deleted.

When I discussed the process of vowel reduction earlier in this chapter, I did not consider a deletion candidate. Since the deletion candidate is now added to the list of candidates, we need to add the anti-deletion constraint MAX. The discussion below will therefore focus on where MAX has to be ranked relative to the other constraints in the partial ranking established for Portuguese vowel reduction (see (22) in §2.7).

I also did not consider the possibility of variation. Therefore I did not discuss the location of the critical cut-off in the constraint ranking. (See Chapter 1 §2.2.3 and Chapter 3 §2.3). Locating the position of the critical cut-off will be another focus of the discussion below. In the rank-ordering model of EVAL, variation can only arise when the critical cut-off point occurs relatively high in the constraint ranking. If it occurred at the

bottom of the hierarchy, all candidates will violate at least one constraint above the cut-off and then no variation will be observed. It is only as the cut-off moves up through the hierarchy that it will reach a point where more than one candidate can be included in the set of candidates disfavored by no constraints above the cut-off. Non-variation is the default situation – most phonological phenomena are categorical. I therefore assume that the critical cut-off point is located as low as possible on the constraint hierarchy.

This is just an extension of the principle of ranking conservatism that I already introduced above in §2.2.1. There I argued that faithfulness rank below markedness constraints by default. Now I am assuming that the critical cut-off ranks below both markedness and faithfulness constraints by default. Constraints are ranked above the cut-off unless if there is positive evidence to the contrary. This positive evidence would take on the following form: If two candidates both appear as outputs for some input, then all constraints that disfavor these two candidates have to rank lower than the cut-off.

3.2 Variation between [ŭ] and ∅

From the table in (24) the following can be calculated about the realization of [ŭ]: (i) *Final in prosodic word*. Silva's data contained 70 instances where [ŭ] could have appeared final in the prosodic word. In 48 of these instances the [ŭ] was deleted. The frequency of the two variants in this context is therefore: ∅ = 69%, [ŭ] = 31%. (ii) *Elsewhere in the prosodic word*. There are 112 instances in the data where [ŭ] could have appeared in a position elsewhere in the prosodic word. In only 16 of these did the [ŭ] delete. The frequency of the variants in this context is therefore: ∅ = 14%, [ŭ] = 86%. Final in the prosodic word the deletion candidate is preferred, while the retention candidate is preferred elsewhere. In prosodic word final contexts, EVAL therefore has to

impose the ordering |∅# ™ ŭ#|[35] on the candidate set, and elsewhere in the prosodic word EVAL has to impose the opposite ordering |ŭ ™ ∅|. I will first discuss the "elsewhere" case, and then argue that the higher deletion rate in prosodic word final position is due to the fact that vowels in this position violate an additional markedness constraint.

3.2.1 Non-final in a prosodic word: |ŭ ™ ∅|

There are three different inputs that can result in an [ŭ] output, namely /ɔ, o, u/ (see §2.3 above). In order for the grammar to impose the ordering |ŭ ™ ∅| on the candidate set, it is necessary that the highest ranked constraint that favors [ŭ] over ∅ dominates the highest ranked constraint that favors ∅ over [ŭ]. The constraints violated by the each of the mappings /o/ → [ŭ], /ɔ/ → [ŭ], /u/ → [ŭ], and /o, ɔ, u/ → ∅ are listed in (25).

(25) **Violation profiles of mappings /o/ → [ŭ], /ɔ/ → [ŭ], /u/ → [ŭ], and /o, ɔ/ → ∅**

/u/ → [ŭ]	*ŏ/u
/o/ → [ŭ]	*ŏ/u, IDENT[-high]
/ɔ/ → [ŭ]	*ŏ/u, IDENT[-high], IDENT[ATR]
/o, ɔ, u/ → ∅	MAX

MAX therefore favors [ŭ] over ∅, while ∅ is favored by *ŏ/u, IDENT[-high] and IDENT[ATR]. The required ranking to ensure the ordering |ŭ ™ ∅| is given (26).

---

[35] In order to distinguish forms that occur in final position in a prosodic word, I will use the symbol # to indicate a prosodic word boundary. A vowel in prosodic word final position will therefore be indicated as /v#/, while a vowel occurring elsewhere in a prosodic word will be indicated simply as /v/.

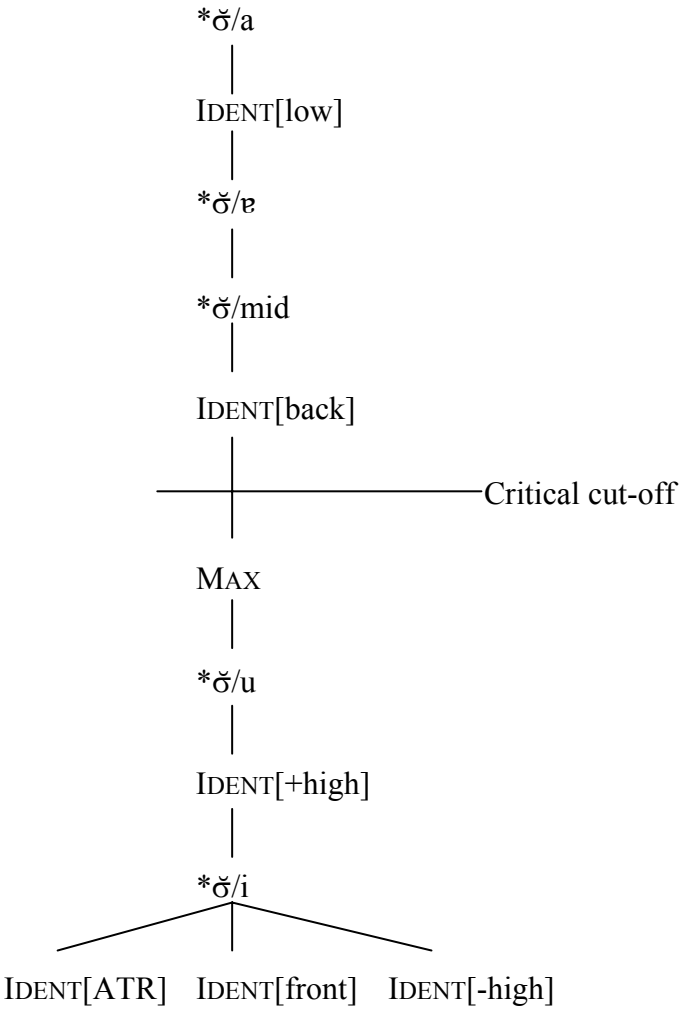
I will also include the symbol # in underlying representations, which is strictly speaking not correct. Prosodic structure is usually assigned by the grammar not part of the input. However, when I use the symbol # in an underlying representation, it should be interpreted as follows: A prosodic word boundary will be inserted by the grammar in this position in the underlying representation.

(26) **Ranking required for |ŭ ™ ∅|**

||MAX o {*ŏ/u, IDENT[-high], IDENT[ATR]}||

Comparison with the ranking in (22) shows that this is compatible with simply adding MAX to a position above *ŏ/u. Following the principle of ranking conservatism, MAX is ranked below all of the markedness constraints ranked higher than *ŏ/u – i.e. ||*ŏ/a o *ŏ/ɐ o *ŏ/mid o MAX||. Adding this information to the ranking from (22) therefore results in the new ranking in (27). In (27) I also indicate the position of the critical cut-off. The motivation for this location of the cut-off follows later in this section.

(27) **Adding MAX to the hierarchy of (22)**



The tableau in (28) shows that the ranking in (27) does indeed result in the rank-ordering |ŭ ™ ∅|. (On the typographical conventions used in this tableau see Chapter 1 §2.2.1 and §2.2.3.)

(28) **/ɔ/→ |ŭ ™ ∅|**

/ɔ/		MAX	*ŏ/u	ID[-hi]	ID[ATR]
2	ŭ		*	*	*
1	∅	*			

Output of EVAL

ŭ *ŏ/u
|
∅ MAX

We still need to find where the critical cut-off point is. This is determined as follows: (i) No candidate that is observed as a variant should be disfavored by any constraint ranked higher than the cut-off. (ii) All candidates that are not observed as variants should be disfavored by at least one constraint ranked higher than the cut-off. As shown just above in (25), the observed variants violate the constraints {MAX, *ŏ/u, IDENT[-high], IDENT[ATR]}. All of these constraints are therefore ranked lower than the cut-off.

Recall the conservative assumption about the location of the cut-off – it is ranked as low as possible (see the discussion at the end of §3.1). Inspection of the hierarchy in (27) will show that this implies that the cut-off is located right between IDENT[back] and MAX, i.e. ||IDENT[back] o Cut-off o MAX||.

(29) **Location of the critical cut-off**

||IDENT[back] o Cut-off o MAX||

Of the candidates that are not observed as variants all but [ŏ, ɔ̆] violate IDENT[back]. With the exception of [ŏ, ɔ̆], all non-observed candidates do therefore violate a constraint ranked higher than the cut-off. The two candidates [ŏ, ɔ̆] both violate *ŏ/mid. Inspection of (27) will show that *ŏ/mid is also ranked higher than the cut-off. The candidates [ŏ, ɔ̆] therefore also violate a constraint ranked higher than the cut-off.

With these refinements to the constraint hierarchy of Portuguese, it is now true that: (i) except for ∅ and [ŭ], all candidates for the inputs /o, ɔ, u/ violate constraints above the critical cut-off; (ii) [ŭ] is rated as more harmonic than ∅. This is shown in the tableau in (30). This tableau considers only an /ɔ/ input, but is representative of the inputs /o, u/ also. The mappings /o, u/ → [ŭ] violate only a subset of the constraints violated by /ɔ/→ [ŭ]. Any ranking that allows the latter will therefore also allow the former.

(30) **/ɔ/→ |ŭ ™ ∅|**

/ɔ/	*ŏ/mid	ID[ba]	MAX	*ŏ/u	ID[-hi]	ID[ATR]
1 ŭ				*	*	*
2 ∅			*			
ŏ, ɔ̆	*!					(*)
all other cands		*!				

Output of EVAL

ŭ *ŏ/u

∅ MAX

Cut-off

··· *IDENT[back] or higher

Neither [ŭ] nor ∅ violates any constraints ranked higher than the cut-off. Both of these candidates will therefore be observed as variants for the input /ɔ/. Because of the ranking ||MAX o *ŏ/u|| EVAL imposes the rank-ordering |ŭ ™ ∅| on these two candidates. From this follows the prediction that [ŭ] will be the more frequent variant of the two. The candidates [ŏ, ɔ̆] violate *ŏ/mid. Since *ŏ/mid is ranked above the cut-off, [ŏ, ɔ̆] will never be accessed as outputs. All other candidates violate at least IDENT[back]. Since IDENT[back] is ranked above the cut-off, these "other candidates" will also never be selected as output. The correct prediction is therefore made: Only [ŭ] and ∅ are observed as variants, and of these two [ŭ] is the more frequent variant.

3.2.2 Final in a prosodic word: |∅# ™ ŭ#|

Unstressed /o#, ɔ#, u#/ that occur in final position in a prosodic word are also variably realized as either ∅# or [ŭ#]. However, in this position ∅# is the more frequent variant. It is therefore necessary that EVAL imposes the rank-ordering |∅# ™ ŭ#| on these two candidates. We know that with the constraints that we have been using up to now, EVAL imposes the opposite rank-ordering on these two candidates. We therefore have to call on an additional constraint. This additional constraint must be violated by [ŭ#] but not by ∅#. I propose that this constraint is a constraint against unstressed vowels in prosodic word final position. Call this constraint *v̆]$_{\omega}$.[36] The violations of the two variants for the inputs /o#, ɔ#, u#/ are listed below in (32).

---

[36] This constraint is probably closely related to final extrametricality (Hayes, 1982, 1995), and it can be restated in terms that will make this connection clearer – something like "do not allow any unfooted vowels" or "do not allow vowels that are not incorporated into prosodic structure".

Evidence for the existence of this (or a very similar) constraint can be found in the process of [ɹ]-intrusion in some dialects of English (Bakovic, 1999, Johansonn, 1973, Kahn, 1976, McCarthy, 1991,

(31) ***v̆]$_\omega$**

A prosodic word is not allowed in to end in an unstressed vowel.

(32) **Violation profiles of mappings /o#/ → [#ŭ], /ɔ#/ → [ŭ#], and /o#, ɔ#/ → ∅#**

/u#/ → [ŭ#]	*v̆]$_\omega$, *ŏ/u
/o#/ → [ŭ#]	*v̆]$_\omega$, *ŏ/u, IDENT[-high]
/ɔ#/ → [ŭ#]	*v̆]$_\omega$, *ŏ/u, IDENT[-high], IDENT[ATR]
/o#, ɔ#, u#/ → ∅#	MAX

In order for EVAL to impose the ordering |∅# ™ ŭ#| on these candidates, it is necessary that the highest ranked constraint that favors ∅# over [ŭ#] dominates the highest ranked constraint that favors [ŭ#] over ∅#. This means that at least one of the

---

1993, Pullam, 1976, Vennemann, 1972). In this phenomenon a vowel final prosodic word is avoided by insertion of an [ɹ], and it results in pronunciations such as *The spa*[ɹ] *is broken* for the sentence *The spa__ is broken*. McCarthy uses a constraint that is very similar to *v̆]$_\omega$ to account for this process (McCarthy, 1993).

Avoidance of final (unstressed) vowels is also attested as a historical change in several languages. Words that ended on /e/ in Old Galician have been reanalyzed lexically in present day Galician without the final /e/ (Martinez-Gil, 1997), so that Old Galician *papele* 'paper' corresponds to Modern Galician *papel*. This is a particularly relevant example, since Galician is closely related to Portuguese.

Several Semitic languages also underwent a process in which final unstressed vowels were deleted. For instance, in Proto-Semitic nouns ended in unstressed vowels /a, i, u/ that indicated the case of the noun. However, these case endings were lost in Hebrew and Aramaic (Moscati, 1964:94-96, O'Leary, 1969:137). This lead to developments such as the following in the word for 'book': Hebrew **sipru* > *sefer*, and Aramaic **sipru* > *səfar* – on the other changes in these words see *inter alia* Coetzee (1996a, 1996b, 1999a, 1999b) and Malone (1972, 1993).

The case endings were retained in Classical/Qur'anic Arabic, but in the dialects of Modern Arabic they have also been deleted (Haywood and Nahmad, 1965:498-499). This lead to differences such as the following between Classical Arabic and modern colloquial dialects: 'house' Classical Arabic *baitu* > modern colloquial Arabic *bait*.

There also seems to be a similar constraint defined on a morphological rather than prosodic domain. McCarthy and Prince (1990a, 1990b) have argued that *stems* in Classical Arabic are not allowed to end in vowels. The constraint FREE-V used by Prince and Smolensky (1993: Chapter 7, no. (152)) in their analysis of Lardil truncation can also be interpreted as a ban on (nominative noun) stems ending in vowels.

constraints violated by the three non-deletion mappings has to dominate MAX. I have argued just above MAX dominates {*ŏ/u, IDENT[-high], IDENT[ATR]}. This leaves only the new constraint *ŭ]$_{\omega}$ to dominate MAX. We therefore need the ranking ||*ŭ]$_{\omega}$ o MAX||. The relation between *ŭ]$_{\omega}$ and the critical cut-off still needs to be determined. Since the non-deletion candidate that does violate *ŭ]$_{\omega}$ is observed as one of the variant outputs, it follows that the cut-off has to be above *ŭ]$_{\omega}$, ||cut-off o *ŭ]$_{\omega}$||. All of the non-observed candidates will still violate either a markedness constraint ranked higher than the cut-off or the faithfulness constraint IDENT[back] which is ranked higher the cut-off. The only observed variants will therefore still be ∅# and [ŭ#].

The rankings required to account for the variation between ∅# and [ŭ#] are summarized in (33). The tableau in (34) shows that these rankings are indeed adequate to explain the variation. In this tableau I consider as before only an /ɔ#/ input – see the discussion above (30) for a motivation.

(33) **Ranking required for |∅# ™ ŭ#|**

||IDENT[back] o cut-off o *ŭ]$_{\omega}$ o MAX||

(34) **/ɔ#/→ |∅# ™ ŭ#|**

/ɔ#/		*ŏ/mid	ID[ba]	*ŭ]$_{\omega}$	MAX	*ŏ/u	ID[-hi]	ID[ATR]
2	ŭ#			*		*	*	*
1	∅#				*			
	ŏ#, ɔ̆#	*!						(*)
	all other cands		*!					

((34) continued)

Output of EVAL

∅# MAX

ŭ# *ŭ)ω

Cut-off

· · · IDENT[back] or higher

The only candidates that violate no constraint ranked higher than the cut-off, are [ŭ#] and ∅#. These two are therefore correctly predicted as the only variants. Because of the ranking ||*ŭ]$_{\omega}$ o MAX||, EVAL imposes the harmonic ordering |∅# ™ ŭ#| on these two candidates. From this follows the prediction that the deletion candidate will be the more frequent variant in prosodic word final position.

3.3 Variation between [ə̆] and ∅

From (24) we can calculate the following about the realization of [ə̆]: (i) *Final in prosodic word*. Silva's data contained 45 instances where [ə̆] could have appeared final in the prosodic word. In 32 of these instances the [ə̆] was actually deleted, i.e. ∅ = 71%, [ə̆] = 29%. (ii) *Elsewhere in the prosodic word*. There are 118 instances in the data where [ə̆] could have appeared in a position elsewhere in the prosodic word, and [ə̆] deleted in 36 of these positions, i.e. ∅ = 31%, [ə̆] = 69%. The same relationship that holds between the retention and the deletion candidate in the [ŭ]~∅ alternation, also holds in the [ə̆]~∅ alternation. Final in a prosodic word, deletion is the preferred variant. Elsewhere in the prosodic word, retention is preferred. In prosodic word final contexts EVAL must impose

the ordering |∅# ™ ə#|, and elsewhere the opposite |ə ™ ∅|. As before, I will begin by considering the elsewhere environment.

3.3.1 Non-final in a prosodic word: |ə ™ ∅|

There are three possible inputs that can result in a [ə] output, namely /e, ɛ, ə/ (see §2.4 and §2.6 above). It is necessary that the highest ranked constraint that favors [ə] over ∅ outranks the highest ranked constraint that favors ∅ over [ə]. In order to determine which constraints favor which of these candidates, I list the constraints violated by each of the following mappings in (35): /e/ → [ə], /ɛ/ → [ə], /ə/ → [ə], and /ə, e, ɛ/ → ∅.

(35) **Violation profiles of mappings /e/ → [ə], /ɛ/ → [ə], /ə/ → [ə], and /ə, e, ɛ/ → ∅**

/ə/ → [ə]	—
/ɛ/ → [ə]	IDENT[front]
/e/ → [ə]	IDENT[front], IDENT[ATR]
/ə, e, ɛ/ → ∅	MAX

MAX is the only constraint that prefers the retention candidate over ∅. In order for retention to be preferred over deletion, it is therefore necessary that MAX ranks above all the constraints violated by the other three mappings. Since we have not yet seen any evidence that required IDENT[front] or IDENT[ATR] to outrank any markedness constraints, both of them are still ranked right at the bottom of the hierarchy, and therefore below MAX. No additional rankings are required for EVAL to impose the ordering |ə ™ ∅| on these two candidates. We also know that the critical cut-off is situated

above MAX, and therefore above all constraints violated by the two variants [ə̆] and ∅. Both of these are therefore predicted as possible outputs.

However, we still need to ascertain that the cut-off point is situated sufficiently low on the hierarchy to eliminate all other vowels as variants – that is, the cut-off should be sufficiently low that all non-observed candidates violate at least one constraint ranked higher than the cut-off. In (36) I list one constraint ranked higher than the cut-off for each of the non-observed candidates (with the exception of [ĭ]). The fact that these constraints are ranked higher than the cut-off can be confirmed by inspecting (27).

(36) **Violations ruling out [ĕ, ɛ̆, ŏ, ɔ̆, ă, ɐ̆, ŭ]**

[ĕ, ɛ̆, ŏ, ɔ̆]	*ŏ/mid
[ă]	*ŏ/a
[ɐ̆]	*ŏ/ɐ
[ŭ]	IDENT[back]

The only non-observed candidate that presents us with a problem is [ĭ]. I propose that [ĭ] is prevented from surfacing as a variant by the local conjunction (Smolensky, 1995) of IDENT[-high] and *ŏ/mid.[37] The idea is that Portuguese is willing to tolerate [ĭ] in an unstressed syllable (/i/ does not reduce), but only if this [ĭ] is the result of a faithful mapping. Portuguese does not tolerate an [ĭ] that is the result of an unfaithful mapping.[38] I define this locally conjoined constraint in (37).

---

[37] For a discussion of the local conjunction of markedness and faithfulness constraints, see Łubowicz (1998, 1999, 2002).

[38] [ĭ] can also be ruled out as an alternative in a model of OT that relies on comparative markedness (McCarthy, 2002a). Portuguese is then willing to violate *ŏ/i_{Old} but not *ŏ/i_{New}. The ranking ||*ŏ/i_{New}

(37) **[IDENT[-high] & *ŏ/i]$_{segment}$**

Do not violate IDENT[-high] and *ŏ/i in the same segment.

If we rank [IDENT[-high] & *ŏ/i]$_{segment}$ higher than the cut-off, then [ĭ] will violate a constraint ranked higher than the cut-off and will therefore be ruled out as possible variant.[39]

In (38) I list the rankings that are required to explain the variation between ∅ and [ə̆]. In (39) I show that this ranking is indeed sufficient. In the tableau I include only those candidates that do not violate a markedness constraint ranked higher than the cut-off. I use /e/ as a representative input. The mapping /e/ → [ə̆] violates a superset of the faithfulness constraints violated by the other two inputs. Any ranking that allows /e/ → [ə̆] will therefore also allow the other two.

(38) **Ranking required for |ə ™ ∅ |**

||[IDENT[-high] & *ŏ/i]$_{segment}$ o Cut-off o MAX o {IDENT[front], IDENT[ATR]}||

---

o Cut-off o *ŏ/i$_{Old}$|| would then account for the pattern. The faithful mapping /i/ → [ĭ] will violate *ŏ/i$_{Old}$ and will surface unproblematically. However, an [ĭ] that is the result of an unfaithful mapping will violate *ŏ/i$_{New}$ and will be ruled out because *ŏ/i$_{New}$ is ranked above the cut-off.

39 The mappings represented by /ə, e, ɛ/ → [ĭ] also violate the constraints *ŏ/i, IDENT[front], IDENT[ATR] and IDENT[-high]. It is therefore at least potentially possible to block these mappings by ranking one of these constraints higher than the cut-off. However, this is for independent reasons not possible.

IDENT[front] and IDENT[ATR] are violated by the mappings /e, ɛ/ → [ə̆], and [ə̆] is one of the variants observed for these inputs. If either of these two constraints ranked higher than the cut-off, then [ə̆] could not surface as variant for these inputs. (See (35) just above.

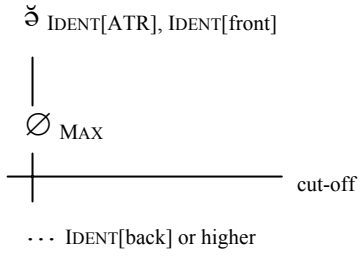
IDENT[-high] is violated by the mappings /o, ɔ/ → [ŭ], and [ŭ] is one of the variants observed for the inputs /o, ɔ/. If IDENT[-high] ranked higher than the cut-off, then [ŭ] could not surface as variant for the inputs /o, ɔ/. (See §3.2.1 (30) above.)

We can also not rank *ŏ/i above the cut-off. In (10) above I argued for the universal ranking ||*ŏ/u o *ŏ/i||. Ranking *ŏ/i above the cut-off will therefore imply by transitivity of constraint ranking that *ŏ/u is also above the cut-off. This will again incorrectly imply that [ŭ] cannot be a variant output for the inputs /o, ɔ, u/. (See §3.2.1 (30) above.)

(39) **/e/→ |ə ™ ∅|**[40]

/e/		*ə̆/mid	[ID[-hi] & *ə̆/i]$_{Seg}$	ID[ba]	MAX	*ə̆/u	*ə̆/i	ID[ATR]	ID[fr]	ID[-hi]
1	ə							*	*	
2	∅				*					
	ĕ	*!								
	ĭ		*!				*			*
	ŭ			*!		*				

Output of EVAL



Only [ə] and ∅ violate no constraints ranked above the cut-off. It is therefore correctly predicted that these will be the only two observed variants. Because of the ranking ||MAX o {IDENT[ATR], IDENT[front]}|| EVAL imposes the harmonic ordering |ə ™ ∅| on these two candidates. It is therefore also correctly predicted that the retention

[40] I am making the following assumption about local conjunctions between markedness and faithfulness constraints with regard to ranking conservatism (see §2.1.1 above): Let [M & F] be such a local conjunction. (i) Since one of the conjuncts of [M & F] is a markedness constraint, I am assuming that it ranks higher than any faithfulness constraint in the absence of evidence to the contrary. (ii) However, since one of the conjuncts of [M & F] is a faithfulness constraint, I assume that it will rank lower than any markedness constraint in the absence of evidence to the contrary. I therefore assume the following default ranking with regard to such local conjunctions: ||Markedness o [M & F] o Faithfulness||.

I am therefore assuming the ranking ||*ə̆/mid o [IDENT[-high] & *ə̆/i]$_{segment}$ o IDENT[back]|| here. This ranking is not necessary to account for the data, and it is assumed simply to make the exposition easier. For more on the decision to follow the principle of ranking conservatism, see §2.1.1 above.

candidate will be the more frequent variant of the two. The faithful candidate of an /e/ or an /ɛ/ input violates *ŏ/mid, which is ranked higher than the cut-off. It is therefore excluded from surfacing as variant. The candidate [ŭ] violates IDENT[back] which is also ranked above the cut-off. This candidate is therefore also not a possible variant. Lastly, consider the candidate [ĭ]. This candidate violates *ŏ/i and IDENT[-high], both of which rank below the cut-off. However, since it violates both of these constraints, it also violates the local conjunction of these two constraints. This locally conjoined constraint ranks higher than the cut-off, so that [ĭ] is prevented from surfacing as a possible variant.

3.3.2 Final in a prosodic word: |∅# ™ ə̆#|

Unstressed /e#, ɛ#, ə#/ are also variably realized as either ∅# or [ə̆#] when they occur in prosodic word final position. However, in this context the deletion candidate is the more frequent variant. EVAL therefore has to impose the following rank-ordering on the candidate set in this context: |∅# ™ ə̆#|. When the variation between ∅# and [ŭ#] was discussed in §3.2.2, I argued for the existence of the constraint *v̆]$_{\omega}$, violated by an unstressed vowel that occurs in prosodic word final position. The mappings /e#, ɛ#, ə#/ → [ə̆#] will also violate this constraint. I have also argued in §3.2.2 (33) that *v̆]$_{\omega}$ is ranked between the cut-off and MAX. This is what explains that the retention candidate that violates *v̆]$_{\omega}$ is less harmonic than the deletion candidate that violates MAX. Finally, in §3.3.1 (38) just above I argued for the locally conjoined constraint [IDENT[-high] & *ŏ/i]$_{segment}$, ranked above the cut-off. This constraint eliminates [ĭ#] as a possible variant. From just these constraints and the ranking already established for them, it follows that: (i) For the inputs /e#, ɛ#, ə#/, ∅# and [ə̆#] are the only candidates that do

not violate any constraints ranked above the cut-off. These candidates are therefore the only two variants. (ii) The deletion candidate ∅# is more harmonic than [ə̆#], and is therefore the more frequently observed of the two variants. This is confirmed by the tableau below in (40). As before, I include only those candidates that do not violate a markedness constraint ranked higher than the cut-off. Also as before, I use /e#/ as a representative input. See the discussion above (38) for a motivation for this move.

(40) **/e#/→ |∅# ™ ə̆#|**

/e#/		*ə̆/mid	[ID[-hi] & *ə̆/i]$_{Seg}$	ID[ba]	*v̆]$_{\omega}$	MAX	*ə̆/u	*ə̆/i	ID[ATR]	ID[fr]	ID[-hi]
2	ə̆#				*				*	*	
1	∅#					*					
	ĕ#	*!			*						
	ĭ#		*!		*			*			*
	ŭ#			*!	*		*				

Output of EVAL

∅# MAX

|

ə̆# *v̆]ω

┼ cut-off

··· IDENT[back] or higher

The only candidates that do not violate a constraint ranked higher than the cut-off are [ə̆#] and ∅#. These are correctly predicted as the only observed variants. Because of the ranking ||*v̆]$_{\omega}$ o MAX|| EVAL imposes the ordering |∅# ™ ə̆#| on these two candidates.

From this follows that the deletion candidate will be the more frequently observed variant in this context. For the inputs /e#, ɛ#/ the faithful candidates will violate *ŏ/mid. Since *ŏ/mid ranks higher than the cut-off, these candidates will not surface as variants. The candidate [ŭ#] violates IDENT[back], which ranks higher than the cut-off. It is also excluded as a possible output. Finally, [ĭ#] violates the locally conjoined [IDENT[-high] & *ŏ/i], which ranks higher the cut-off, and therefore eliminates [ĭ#] as a variant.

3.4 Interim summary

In the table in (41) I summarize all of the new rankings that have been argued for since the last summary in (21) in §2.7. As in (21) the first column lists the ranking, the second states the motivation for the ranking, and the last column states where in the text the ranking was introduced. In (42) I give a graphic representation of the Faialense Portuguese constraint hierarchy as it stands at this point in the discussion. In this graphic representation I do not include the hierarchy above *ŏ/mid. We have not encountered any evidence prompting a change in this part of the hierarchy since the last summary in (21). This part of the hierarchy therefore still looks exactly the same as in (21).

(41) **Summary of rankings thus far**

Ranking	**Motivation**	**Where?**
MAX ○ *ŏ/u, ID[-hi], ID[ATR]	More [ŭ] that ∅ for /o, ɔ, u/[41]	§3.2.1 (26) (28)
ID[ba] ○ Cut-off ○ MAX	[ŭ] and ∅ only variants for /o, ɔ, u/[42]	§3.2.1 (29) (30)

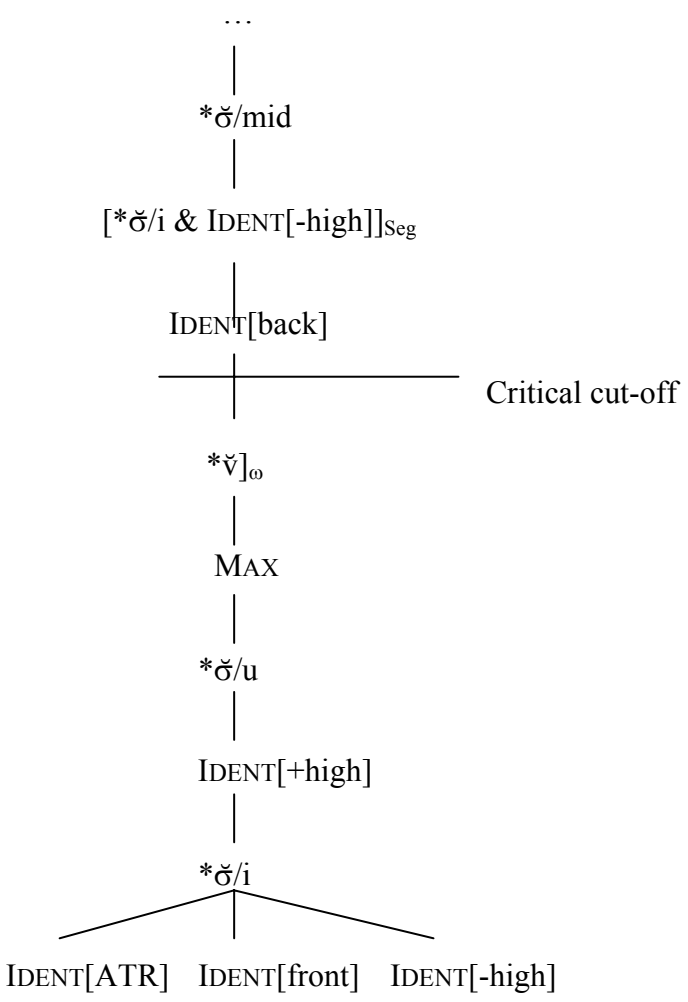
[41] The ranking argument used to motivate this ranking is different from the ranking arguments used in classic OT. The argument is not that the opposite ranking will result in selecting the wrong output. The argument is rather that the opposite ranking will result in the wrong relative frequency relation between the variants. See Chapter 3 §2.4 for a discussion of this.

[42] See previous footnote.

((41) continued)

Ranking	**Motivation**	**Where?**
Cut-off o*v̆]$_{\omega}$, MAX	∅# and [ŭ#] only variants for /o#, ɔ#, u#/	§3.2.2 (33) (34)
[ID[-hi] & *ŏ/i]$_{Seg}$ o Cut-off	[ĭ] not observed as variant for /e, ɛ, ə/	§3.3.1 (38) (39)
MAX o ID[fr], ID[ATR]	more [ə̆] than ∅ for /e, ɛ, ə/	§3.3.1 (38) (39)
*ŏ/mid o [*ŏ/i & ID[-hi]]$_{Seg}$ o ID[ba]	Ranking conservatism	§3.3.1 Footnote 40.

(42) **Graphic representation of the Portuguese hierarchy established thus far**



3.5 Variation between [ĭ] and ∅

Because of the stress placement rules of Portuguese, [ĭ] cannot appear final in a prosodic word (see footnote 8 in §1.2 and footnote 29 in §3.1). We therefore only have to account for the deletion pattern of [ĭ] that occurs in non-final position in the prosodic word. From the table in (24) we see that Silva's corpus contains 75 instances of [ĭ] that could have appeared in unstressed syllables in non-final position in a prosodic word. Of these 5 were deleted, i.e. [ĭ] = 93%, ∅ = 7%. The retention candidate is preferred over the deletion candidate, and we therefore need EVAL to impose the rank-ordering |ĭ ™ ∅| on these two candidates.

There is only one input that can result in an [ĭ] output, namely /i/. This is because no vowel reduces to [ĭ] – see §1.2 (6) above. In order for EVAL to impose the rank-ordering |ĭ ™ ∅| on the candidate set, it is necessary that the highest ranked constraint that favors [ĭ] over ∅ be ranked higher than the highest ranked constraint that favors ∅ over [ĭ]. I list the violations that each of the two observed mappings earns in (43).

(43) **Violation profiles of the mappings /i/ → [ĭ], and /i/ → ∅**

/i/ → [ĭ]	*ŏ/i
/i/ → ∅	MAX

Since MAX is the only constraint that favors [ĭ] over ∅, MAX has to outrank all constraints violated by [ĭ], i.e. ||MAX o *ŏ/i||. Comparison with (42) above shows that this is consistent with the rankings established thus far. Since these two candidates are

observed as variants, it is also necessary that neither of them violate any constraints ranked higher than the cut-off. Again, comparison with (42) shows that both MAX and *ɵ̆/i are ranked lower than the cut-off. It therefore follows that [ɨ̆] and ∅ will be observed as variants, and that [ɨ̆] will be the more frequent variant of the two.

All that still needs to be determined is where the cut-off has to be ranked to assure that no other candidate will surface as a variant. Each of the non-observed candidates has to violate at least one constraint ranked higher than the cut-off. With the exception of [ə̆] each of the non-observed variants violate a constraint that is already ranked above the cut-off. In (44) I list one constraint ranked higher than the cut-off for each of the non-observed candidates (with the exception of [ə̆]). The fact that these constraints are ranked higher than the cut-off can be confirmed by inspecting (42) and (27).

(44) **Violations ruling out [ĕ, ɛ̆, ŏ, ɔ̆, ă, ɐ̆, ŭ]**

[ĕ, ɛ̆, ŏ, ɔ̆]	*ɵ̆/mid
[ă]	*ɵ̆/a
[ɐ̆]	*ɵ̆/ɐ
[ŭ]	IDENT[back]

The mapping /i/ → [ə̆] violates IDENT[+high], IDENT[ATR] and IDENT[front]. If any of these three constraints ranked higher than the cut-off, then we can prevent [ə̆] from surfacing as variant output for /i/. Of these three constraints, only IDENT[+high] can be ranked higher than the cut-off. The reason is that IDENT[ATR] and IDENT[front] are both violated by one of the observed variants for an /e/-input. An /e/-input is variably realized as either ∅ or [ə̆] (see §3.3.1 above). The mapping /e/ → [ə̆] violates both IDENT[ATR]

and IDENT[front]. If either of these constraints were to rank higher than the cut-off, then [ə̆] would be blocked from surfacing as variant output for /e/. We can therefore not use IDENT[ATR] or IDENT[front] to prevent [ə̆] from surfacing as variant output for /i/.

IDENT[+high], on the other hand, is not violated by a variant output for any input. There are only two inputs that could violate IDENT[+high], namely /i, u/. Neither of these inputs has an observed variant that violates IDENT[+high]. There is therefore nothing that prevents us from ranking IDENT[+high] above the cut-off. Since IDENT[+high] is a faithfulness constraint, it will be ranked as low as possible with respect to markedness constraints – following the principle of ranking conservatism (see §2.1.1 and footnote 40 in §3.3.1 above). It will therefore be ranked below all of the markedness constraints (and markedness + faithfulness conjunctions) above the cut-off. In particular, this means that it will rank below [IDENT[-high] & *ŏ/mid]$_{\text{Segment}}$. The rankings required to account for the variation between [ĭ] and ∅ are stated in (45).

(45) **Ranking required for |ĭ ™ ∅|**

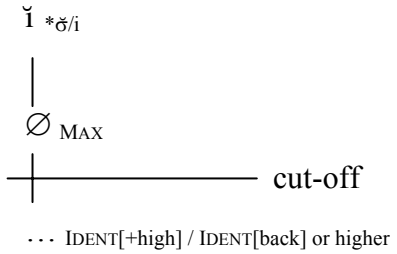
||[IDENT[-high] & *ŏ/mid]$_{\text{Segment}}$ o IDENT[+high] o cut-off o MAX o *ŏ/i||

With these rankings: (i) [ĭ] and ∅ are the only candidates that violate no constraints ranked higher than the cut-off. They are therefore predicted as the only two observed outputs for an /i/ input. (ii) [ĭ] is rated better than ∅, and retention is therefore predicted to be preferred over deletion. This is confirmed by the tableau in (46). In this tableau I do not consider any of the vowels that violate markedness constraints higher than the cut-off.

(46) **/i/→ [ĭ ™ ∅]**

/i/		ID[ba]	ID[+hi]	MAX	*ŏ/u	*ŏ/i	ID[ATR]	ID[fr]
1	ĭ					*		
2	∅			*				
	ə̆		*!				*	*
	ŭ	*!			*			*

Output of EVAL



The only two candidates that are not disfavored by any constraint ranked higher than the cut-off are [ĭ] and ∅. These candidates are therefore correctly predicted as the only two variants for an /i/ input. Because of the ranking ||MAX o *ŏ/i||, EVAL imposes the rank-ordering [ĭ ™ ∅] on these candidates. From this follows the correct prediction that the retention candidate will be the more frequent variant. [ə̆] violates IDENT[+high] and [ŭ] violates IDENT[back]. Both of these constraints are ranked higher than the cut-off and these candidates are therefore blocked from being possible variants. The vowels not considered in (46) violate *ŏ/mid, *ŏ/a or *ŏ/ɐ, all of which also rank higher than the cut-off.

3.6 No variation: always [ɐ̆]

The vowel [ɐ̆] has a very low overall deletion rate. As can be seen in table (24), Silva's corpus contains 331 sites where an [ɐ̆] could have occurred, and in only 7 of these the vowel was deleted. This represents a total deletion rate of just over 2%.[43] Because the deletion rate of [ɐ̆] is so low, I will treat this vowel as if it categorically resists deletion.

When I discussed the critical cut-off in Chapter 1 §2.2.3 I explained that there are two situations in which no variation will arise in the rank-ordering model of EVAL. I repeat these two in (47).

(47) **No variation in a rank-ordering model of EVAL**

a. All candidates except for one violate at least one constraint above the cut-off. The single candidate that does not violate a constraint above the cut-off is then selected as the only "variant".

b. All candidates violate at least one constraint above the cut-off. In such a circumstance the single best candidate is chosen as output even though it does violate a constraint above the cut-off.

The observed output under consideration here is [ɐ̆]. This vowel violates the markedness constraint *ɐ̆/ɐ. Inspection of (42) and (22) will show that *ɐ̆/ɐ is ranked above the cut-off. This means that we are dealing here with the second of the scenarios in (47) above.

---

[43] Since the number of deletions is so small, it is not possible to compare [ɐ̆] that occurs final in the prosodic word and [ɐ̆] that occurs elsewhere in the prosodic word.

There are two inputs that are realized as [ɐ̆] in unstressed syllables, namely /a, ɐ/ (see §1.2 above). In order to ensure that [ɐ̆] is the only observed output for these two inputs, the following is necessary: (i) The highest ranked constraint that favors [ɐ̆] over any other candidate outranks the highest ranked constraint that favors this other candidate over [ɐ̆]. (ii) All other candidates have to violate at least one constraint above the cut-off.[44] The violations incurred by each of the mappings /a/ → [ɐ̆] and /ɐ/ → [ɐ̆] are listed in (48).

(48) **Violation profiles of the mappings /a/ → [ɐ̆] and /ɐ/ → [ɐ̆]**

/a/ → [ɐ̆]	*ɐ̆/ɐ, IDENT[ATR]
/ɐ/ → [ɐ̆]	*ɐ̆/ɐ

Inspection of (22) and (42) will show that of the constraints *ɐ̆/ɐ and IDENT[ATR], *ɐ̆/ɐ ranks the highest. The highest ranked constraint that disfavors [ɐ̆] is therefore *ɐ̆/ɐ. This implies that it is crucial that all other candidates violate at least one constraint ranked higher than *ɐ̆/ɐ. This is not a problem for the other vowels. The other low vowel [ă] violates the constraint *ɐ̆/a, which universally outranks *ɐ̆/ɐ – since /a/ is higher in sonority than /ɐ/ (see (10) above). All other vowels violate the faithfulness constraint IDENT[low]. I have argued above for the ranking ||IDENT[low] o*ɐ̆/ɐ|| (see §2.2 (15) and (16)). This therefore eliminates all vowels except for [ɐ̆]. Only the deletion candidate still needs to be eliminated as possible output. Of the constraints that we have considered up to now, the deletion candidate violates only MAX. However, we cannot use

[44] This second requirement actually follows from the first, since [ɐ̆] does violate *ɐ̆/ɐ, which is ranked above the cut-off.

MAX to block the deletion candidate from surfacing as a variant output. The reason for this is that the deletion candidate does alternate as variant with [ŭ], [ĭ] and [ə̆]. If MAX ranked higher than the cut-off, the deletion candidate would wrongly be prevented from surfacing as variant in these contexts. We therefore need a different constraint to block the deletion candidate from surfacing as variant of [ɐ̆]. I propose that the constraint that is responsible for this is a MAX constraint indexed specifically to the low vowels.[45] This constraint is defined in (49).

(49) **MAX-a/ɐ**

/a, ɐ/ must have some output correspondent.

If MAX-a/ɐ is ranked above *ŏ/ɐ,[46] then ∅ will violate a constraint ranked higher than the constraint violated by the actually observed output [ɐ̆]. This will then result in a situation where [ɐ̆] is the only predicted output. The crucial rankings to ensure that /a, ɐ/ maps only onto [ɐ̆] are given in (50). The tableau in (51) shows that this ranking does indeed result in selecting only [ɐ̆] as output. In this tableau the input /a/ is used as example. Since this input violates a superset of the violations of /ɐ/, any ranking that will allow the mapping /a/ → [ɐ̆], will also allow the mapping the /ɐ/ → [ɐ̆].

---

45 See Hartkemeyer (2000) and Tranel (1999) for arguments in favor of MAX constraints indexed to specific vowels. Their idea is that vowels of higher sonority are protected by higher ranking MAX-constraints than vowels of lower sonority. Since the low vowels /a, ɐ/ are the highest in sonority of the Portuguese vowels (see §1.2 (8) above), the MAX-constraint for these vowels ranks higher than the MAX constraints for the other vowels. The MAX-constraints for the other vowels can therefore be ranked lower than the cut-off. In fact, it is possible to replace the general MAX that I currently use with the MAX-constraints indexed to the non-low vowels. See Gouskova (2003:240) for an argument against vowel-specific MAX-constraints.

46 Since MAX-a/ɐ is a faithfulness constraint, I am ranking it as low as possible in accordance with the conservative ranking principle (§2.1.1). Since we have no evidence that MAX-a/ɐ should rank above *ŏ/a, I am ranking MAX-a/ɐ between *ŏ/a and *ŏ/ɐ.

(50) **Ranking requirement for /a, ɐ/ → [ɐ̆]**

||*ɔ̆/a o {MAX-a/ɐ, IDENT[low]} o *ɔ̆/ɐ o IDENT[ATR]||

(51) **/a/ → [ɐ̆]**

/a/	*ɔ̆/a	MAX-a/ɐ	ID[low]	*ɔ̆/ɐ	MAX	ID[ATR]
1 ɐ̆				*		*
2 ă	*!					
∅		*!			*	
All other cands			*!			(*)

Output of EVAL

——————— cut-off

ɐ̆ *ɔ̆/ɐ

|

. . . *ɔ̆/ɐ or higher

All candidates are disfavored by at least one constraint ranked higher than the cut-off. When this happens, only the single best candidate will be selected as output. The candidate [ɐ̆] violates *ɔ̆/ɐ. All non-low vowels violate IDENT[low] which is ranked higher than *ɔ̆/ɐ. The other low vowel [ă] violates *ɔ̆/a which is also ranked higher than *ɔ̆/ɐ. The deletion candidate violates MAX-a/ɐ which also dominates *ɔ̆/ɐ. It is therefore correctly predicted that [ɐ̆] will be the only observed output.

I am not discussing [ɐ̆] that occurs in prosodic word final position separately. There are two reasons for this. First, [ɐ̆] resists deletion, irrespective of where it occurs in the prosodic word. Secondly, inspection of tableau (51) shows that the same pattern is indeed predicted for the inputs /a#, ɐ#/. The only difference between tableau (51) and one

with /a#/ as input, is that all of the non-deletion candidates will receive a violation in terms of *v̆]$_{\omega}$. However, all candidates already violate a constraint that is ranked higher than *v̆]$_{\omega}$. This violation therefore will have no influence on the selection of the output.

3.7 Final summary

In the table in (52) I summarize all of the new rankings that have been argued for since the last summary in (41) in §3.4. As before, I also give a graphic representation of the Faialense Portuguese constraint hierarchy in (53). This hierarchy represents the grammar that is required to account for vowel reduction and deletion in unstressed syllables in Portuguese.

(52) **Summary of rankings thus far**

Ranking	**Motivation**	**Where?**
ID[+hi] o Cut-off	To exclude [ə̆] as a possible variant of /i/	§3.5 (45) (46)
[ID[-hi] & *ŏ/i]$_{Seg}$ o ID[+hi]	Ranking conservatism	§3.5 (45) (46)
MAX-a/ɐ o *ŏ/ɐ	To exclude ∅ as possible variant for /a, ɐ/	§3.6 (50) (51)
*ŏ/ɐ o MAX-a/ɐ	Ranking conservatism	§3.6 (50) (51)

This concludes the basic account of Portuguese vowel reduction and deletion. I have argued that:

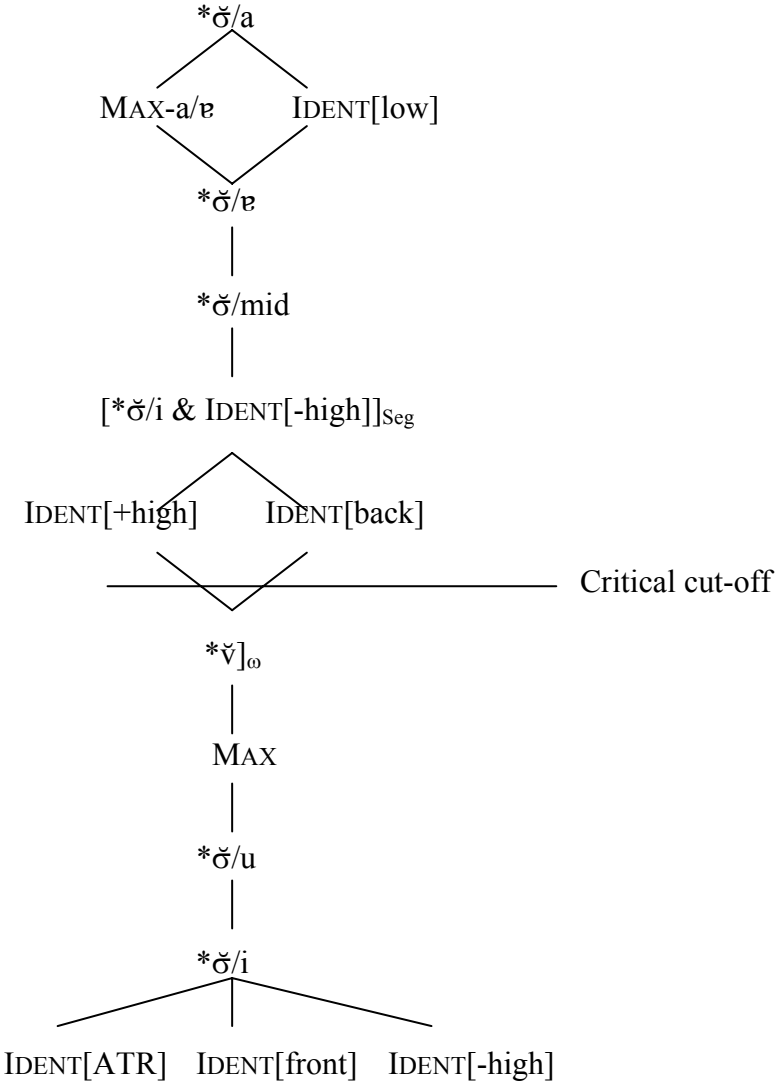
(i) EVAL rank-orders the full candidate set, and does not simply distinguish between the best candidate and the mass of losers.

(iii) There is a critical cut-off in the constraint hierarchy. Candidates that violate constraints ranked above the cut-off, are eliminated as possible variants.

(iv) When all candidates violate constraints above the cut-off, then only the single best candidate is selected as output (= a categorical process).

With these assumptions I was able to present a coherent account of the vowel reduction, and variable vowel deletion in Faialense Portuguese. In the rest of this chapter I will show how we can also account for variation patterns across different inputs by allowing EVAL to compare candidates that are not related via the same input.

(53) **Graphic representation of the Portuguese constraint hierarchy**



4. Variation across contexts

In the previous section I presented an account of why variation is encountered in the realization of individual vowels. In that section I dealt with questions such as: Why is the input /o#/ sometimes realized as [ŭ#] and sometimes as ∅#, and in particular, why is it more often realized as ∅#? This is what I refer to as intra-contextual variation (see Chapter 1 §2.2.1). However, there is also the inter-contextual variation that need to be accounted for (see Chapter 1 §2.2.2). It is often the case that a variable process applies more frequently in one context than another. For instance, for both /i/ and /e, ɛ, ə/ in non-prosodic word final position the retention candidate is the more frequent variant (/i/ → |ĭ ™ ∅| and /e, ɛ, ə/ → |ə̆ ™ ∅|). This fact has been accounted for in §3 above. However, deletion is observed much more frequently for /e, ɛ, ə/ (31%) than for /i/ (7%). There are more such patterns that can only be identified by generalizing over different input forms. To account for these patterns, it is necessary to compare candidates that are not related to each other via the same input, i.e. by considering non-generated comparison sets (see Chapter 1 §1.2). In this section I offer an account for these cross-input patterns by appealing to such non-generated comparison sets.

Silva coded the data in his corpus for several linguistic features, and then subjected the data to analysis within the variable rule framework of Labov (Cedergren and Sankoff, 1974, Kay and McDaniel, 1979, Labov, 1972). In particular, he employs the VARBUL software package (Sankoff and Rand, 1988) to determine which grammatical factors contribute significantly towards determining the observed pattern of deletion. Only three of the factors for which Silva coded his data were selected as significant. I have already discussed these three factors when I initially presented Silva's data on

deletion in §3.1. I briefly mention each of them again. However, for more detailed discussion, refer to §3.1 and the table in (24).

(i) *Vowel quality*. Some vowels are more susceptible to deletion than others. In particular, the vowels can be ordered as follows with regard to how likely they are to delete: [ə̆] (42%) > [ŭ] (35%) > [ĭ] (7%) > [ɐ̆] (2%).[47] (ii) *Position in the prosodic word.* For all inputs, deletion applies more frequently in prosodic word final position (70%), than elsewhere in the prosodic word (19%). (iii) *Stress of the following syllable*. For all inputs, a vowel is more likely to delete if followed by an unstressed syllable (48%) than if followed by a stressed syllable (25%). In the rest of this section I will discuss each of these three observations.

4.1 Vowel quality

There are three forces that co-determine how likely a specific process is to apply: (i) *Markedness of the input*. The more marked the input is, the more likely it is that a process will apply to decrease its markedness. (ii) *Cost of application of the process*. In OT terms application of a process translates into violation of a faithfulness constraint. The faithfulness constraints therefore militate against application of a process. The more faithfulness constraints violated by application of a process, the less likely that process is to apply. Similarly, the higher the ranking of the faithfulness constraint(s) violated by the application of some process, the less likely that process is to apply. (iii) *Markedness of the output*. The more marked the form that results from the application of the process, the less likely the process is to apply.

---

[47] Strictly speaking, it is of course not the surface vowels [ə̆, ŭ, ĭ, ɐ̆] that delete, but their underlying correspondents. It would be most correct to replace each of the vowels in this statement with its possible underlying correspondents as follows: [ə̆] = /e, ɛ, ə/, [ŭ] = /o, ɔ, u/, [ĭ] = /i/, [ɐ̆] = /a, ɐ/.

In order to determine how likely a process is to apply, there are therefore two relevant non-generated comparison sets to consider. First, the non-generated comparison set comprising of candidates, from each of the relevant classes, in which *the process has not applied* ("non-undergoers" of the process). By comparing these forms, it can be determined how they relate to each other in terms of markedness. The prediction is that the most marked of these candidates corresponds to the context where the process is most likely to apply.

Secondly, we need to consider the non-generated comparison set that contain candidates in which *the process has applied* ("undergoers" of the process). We can compare these candidates in terms of their markedness and faithfulness violations. The least marked amongst these will correspond to the context in which the process is most likely to apply. Similarly, from among these candidates that one that does best on the faithfulness constraints, is predicted to correspond to the context in which the process applies most frequently. I consider each of these two non-generated comparison sets below.

4.1.1 The non-undergoers

For each of the possible vowel inputs, the set of non-undergoers will contain a candidate in which the vowel has not deleted. For vowels that are not subject to reduction, the candidate will be the faithful candidate. However, for vowels that are subject to reduction, the candidate will be the reduced vowel.

(54) **The non-generated comparison set of non-undergoers**

Non-undergoers = { /i/ → [ĭ];

/a/ → [ɐ̆]; /ɐ/ → [ɐ̆];

/ɔ/ → [ŭ]; /o/ → [ŭ]; /u/ → [ŭ];

/ɛ/ → [ə̆]; /e/ → [ə̆]; /ə/ → [ə̆] }

Consider first the mappings /a, ɐ/ → [ɐ̆]. These mappings violate the constraint *ŏ/ɐ which is ranked higher than the markedness constraint violated by each of the other mappings. The drive to delete is therefore predicted to be the strongest in this context, and we expect the highest deletion rate associated with this context. However, this is not correct. In fact, the vowel [ɐ̆] has such a low rate of deletion that it was treated above as if it is categorically retained (see §3.6). The prediction here is counter to the facts. However, recall how the deletion candidate was ruled out in for the low vowels. There is faithfulness constraint that specifically militates against deletion of the vowels /a, ɐ/, namely MAX-a/ɐ, and this faithfulness constraint is ranked higher than the critical cut-off. Although the candidate [ɐ̆] is the most marked of all the retention candidates, and although the drive to delete is therefore strongest for this vowel, deletion is categorically blocked by a special faithfulness constraint. In the rest of this section I will not consider [ɐ̆] any further.

Now consider the mappings /ɔ, o, u/ → [ŭ] and /i/ → [ĭ]. The mappings /ɔ, o, u/ → [ŭ] violate the markedness constraint *ŏ/u, while /i/ → [ĭ] violates *ŏ/i. Since /u/ is more sonorous than /i/, *ŏ/u universally outranks *ŏ/i (see (9) and (10) in §2.1). The mappings /ɔ, o, u/ → [ŭ] therefore violate a higher ranked constraint than the mapping

/i/ → [ĭ], and it is predicted that the drive to delete in the [ŭ]-context should be stronger than the drive to delete in the [ĭ]-context. This corresponds to the observed deletion pattern.

Now consider the mappings /ɔ, o, u/ → [ŭ] and /ɛ, e, ə/ → [ə̆]. The mappings represented by /ɔ, o, u/ → [ŭ] violate the markedness constraint *ŏ/u. However, since schwa is the most harmonic unstressed syllable, there is no constraint *ŏ/ə – see §2.1 (9) above and Gouskova (2003). The mappings /ɛ, e, ə/ → [ə̆] therefore violate none of the markedness constraints considered thus far. According to this, there is no markedness constraint that would drive deletion in the [ə̆]-context, while the constraint *ŏ/u drives it in the [ŭ]-context. A higher deletion rate is therefore expected in the [ŭ]-context than in the [ə̆]-context. However, this is counter to the facts. [ə̆] is associated with a higher deletion rate than [ŭ]. In fact, [ə̆] has the highest deletion rate of all vowels in Faialense Portuguese.

If there truly were no markedness constraint violated by [ə̆], then it would be extremely hard to explain why it would ever delete. I propose that [ə̆] violates a general constraint against associating a syllabic nucleus with schwa – i.e. irrespective of whether the relevant syllable is stressed or not. This is just a member of the peak-affinity constraint family of Prince and Smolensky (1993:129). Their argument for the existence of this constraint family can be stated as follows: The syllabic peak (nucleus) is more prominent than the syllabic margins (onset/coda). More sonorous segments are more prominent than less sonorous segments. Prominent elements prefer to co-occur, and therefore a syllabic peak is more well-formed the more sonorous the segment is that is

associated to the syllabic peak. This is again an example of harmonic alignment – see (9) and (10) in §2.1 above. In fact, this is one of the examples that Prince and Smolensky used when they introduced the concept of harmonic alignment. The harmonic alignment of the peak/margin scale and the sonority scale can be represented as in (55).

(55) **Harmonic alignment of peak/margin and sonority**

Syllabic position

prominence: Peak > Margin

Sonority scale: a > ɐ > {o,ɔ,e,ɛ} > u > i > ə > {y,w} > {r,l} … > {t, p, k}

Harmonic alignment

on Peaks: P/a ™ P/ɐ ™ P/{o, ɔ, e, ɛ} ™ P/u ™ P/i ™ P/ə … ™ P/{t, p, k}

Constraint hierarchy: ||*P/{t, p, k} o … *P/ə o *P/i o *P/u o … *P/ɐ||

Here the sonority scale is harmonically aligned not with a prominence scale on different nuclei (more prominent = stressed nuclei vs. less prominent = unstressed nuclei). The sonority scale is aligned with a scale in which all peaks qualify as prominent. The constraints that are formulated based on this harmonic alignment therefore do not distinguish between stressed and unstressed syllables. These constraints are violated by a vowel irrespective of whether it occurs in a stressed or an unstressed syllable.

Since schwa is the least sonorous vowel, the constraint from this family that penalizes schwa (*P/ə) is ranked higher than the constraint that penalizes any of the other vowels. This is the constraint that is responsible for driving deletion of [ə̆] in Faialense Portuguese. Where should *P/ə be ranked? Since [ə̆] is associated with higher deletion

rates than [ŭ], it is our expectation that [ə̆] should be more marked than [ŭ]. Therefore, the constraint violated by [ə̆], should rank higher than the constraint violated by [ŭ], i.e. ||*P/ə o *ŏ/u||.[48] With this additional ranking, we can now account for the relative deletion rates associated with each of the [ŭ], [ə̆] and [ĭ]. This is shown in the tableau in (56). In this tableau I only include the constraints actually violated by at least one candidate.

(56) **Comparing non-deletion candidates from different inputs**

		*P/ə	*ŏ/u	*ŏ/i	ID[ATR]	ID[fr]	ID[-hi]
1	/i/ → [ĭ]			*			
2	/ɔ/ → [ŭ]		*		*		*
2	/o/ → [ŭ]		*				*
2	/u/ → [ŭ]		*				
3	/ɛ/ → [ə̆]	*				*	
3	/e/ → [ə̆]	*			*	*	
3	/ə/ → [ə̆]	*					

Output of EVAL

/i/ → [ĭ] $_{*ŏ/i}$

|

/ɔ, o, u/ → [ŭ] $_{*ŏ/u}$

|

/ɛ, e, ə/ → [ə̆] $_{*P/ə}$

In this tableau I simplify somewhat by ignoring the relationships between the different candidates in the last two groups. However, the general pattern is clear. The mapping /i/ → [ĭ] is preferred over all of the mappings onto [ŭ], which is again preferred

[48] See discussion just below in §4.1.1.1 about the relation between MAX and *P/ə.

over all of the mappings onto [ə̆]. Since the mapping /i/ → [ĭ] is most preferred (= least marked), the drive to delete is the lowest here and we predict the lowest deletion rate associated with /i/. This is indeed correct. Since the mappings /ɛ, e, ə/ → [ə̆] is least preferred (= most marked), the drive to delete is strongest for the inputs /ɛ, e, ə/. We therefore correctly predict the highest deletion rate for these inputs.

By allowing EVAL to compare forms from different inputs, we can capture the intuition that certain forms are more marked and therefore more likely to undergo some process. We can order the three vowels [ĭ], [ə̆] and [ŭ] as follows in terms of their markedness: |ĭ ™ ŭ ™ ə̆|. This explain why [ə̆] deletes most, then [ŭ], and then [ĭ].

4.1.1.1 The peak-affinity constraints in general

In the previous section I argued for the inclusion of a member of the peak-affinity constraints into the set of constraints active in vowel deletion process of Portuguese. It is now necessary to answer two questions about these constraints: (i) Does the introduction of *P/ə influence any of the predictions with regard to reduction and deletion of /e, ɛ, ə/ as discussed above in §3.3? (ii) What about the other members of the peak-affinity hierarchy?

Consider first the influence of *P/ə on the realization of the inputs /e, ɛ, ə/. When any of these vowels are parsed into an unstressed syllable that is not final in a prosodic word, variation between [ə̆] and ∅ is observed with [ə̆] as the most frequent variant. It is therefore necessary that EVAL imposes the following rank-ordering on these two candidates: |ə̆ ™ ∅|. For this to happen, to highest ranked constraint that favors ∅ over [ə̆]

must rank below the highest ranked constraint that favors [ə̆] over ∅. The violation profiles of the mappings /e, ɛ, ə/ → [ə̆] and /e, ɛ, ə/ → ∅ are listed in (57).

(57) **Violation profiles of the mappings /e, ɛ, ə/ → [ə̆] and /e, ɛ, ə/ → ∅**

/ə/ → [ə̆]	*P/ə
/ɛ/ → [ə̆]	*P/ə, IDENT[front]
/e/ → [ə̆]	*P/ə, IDENT[front], IDENT[ATR]
/e, ɛ, ə/ → ∅	MAX

MAX is the only constraint that favors the retention candidates over the deletion candidate. MAX must therefore outrank all of the constraints violated by the retention candidates. The ranking ||MAX o {IDENT[front], IDENT[ATR]}|| has already been established (see (53) in §3.7). All that we need to add now is the ranking ||MAX o *P/ə||. Together with ||*P/ə o *ŏ/u|| argued for just above in 4.1.1, we therefore know exactly where *P/ə ranks, namely: ||MAX o*P/ə o *ŏ/u||. With this ranking *P/ə will not interfere with the predictions that the theory makes about the realization of the inputs /e, ɛ, ə/.

Now we can consider the rest of the peak-affinity hierarchy. Can the other *P/-constraints cause problems? Only if they rank too high. In particular, if the *P/-constraint for some vowel ranks higher than the *ŏ/-constraint for that same vowel. But since *P/ə ranks highest of all the *P/-constraints, we can claim that the other *P/-constraints rank lower than *P/ə, and in particular low enough that they will not have an influence on the observed output patterns.

4.1.2 The undergoers

We still need to consider the set of undergoers – i.e. a candidate for each possible output where the vowel was indeed deleted. All deletion candidates violate the constraint MAX. The inputs /a, ɐ/ violate, in addition to the general MAX constraint, also the special constraint MAX-a/ɐ. The tableau in (58) compares the members of the non-generated comparison set of undergoers. Since MAX and MAX-a/ɐ are the only constraints violated by these candidates, only these two constraints are included in the tableau.

(58) **Comparing deletion candidates from different inputs**

		MAX-a/ɐ	MAX
1	/i/ → ∅		*
1	/ɔ/ → ∅		*
1	/o/ → ∅		*
1	/u/ → ∅		*
1	/ɛ/ → ∅		*
1	/e/ → ∅		*
1	/ə/ → ∅		*
2	/a/ → ∅	*	*
2	/ɐ/ → ∅	*	*

Output of EVAL

/i, ɔ, o, ɛ, e, ə/ → ∅

|

/a, ɐ/ → ∅ MAX-a/ɐ

What this comparison shows is that deletion of /a, ɐ/ comes at a higher faithfulness cost than deletion of /i, ɔ, o, ɛ, e, ə/. Faithfulness will therefore militate stronger against the deletion of /a, ɐ/, with the result that we expect lower deletion rates associated with these two inputs than with the other vowels. This is confirmed by the data

– in fact, /a, ɐ/ deletes so infrequently that I have claimed above that MAX-a/ɐ ranks above the critical cut-off, thereby ruling out deletion as a variant for these inputs.

4.2 Position in the prosodic word

Averaging across all inputs, it can be seen that deletion applies more in prosodic word final position (70%) than elsewhere in the prosodic word (19%). Inspection of the table in (24) shows that this pattern is also individually true for the vowels that are realized as [ə̆] (/e, ɛ, ə/) and as [ŭ] (/o, ɔ, u/). The vowel [ĭ] does not occur in prosodic word final position (footnote 8 §1.2 and footnote 29 §3.1), and we can therefore not compare the deletion rate for this vowel in the two different positions in the prosodic word. For the vowels realized as [ɐ̆] (/a, ɐ/), this generalization does not hold. However, /a, ɐ/ in general resists deletion, irrespective of whether it occurs final or elsewhere in the prosodic word. In the rest of this section, I will therefore deal only with the inputs /e, ɛ, ə/ and /o, ɔ, u/.

We can again consider the set of deletion undergoing mappings and the set non-undergoing mappings as non-generated comparison sets. However, in this instance the comparison between the undergoers will be uninformative. All of the mappings represented by /e, ɛ, ə, o, ɔ, u/ → ∅ violate only one constraint, namely MAX, irrespective of whether the deletion occurs in prosodic word final position or elsewhere in the prosodic word. The faithfulness cost of deletion is the same everywhere in the prosodic word. The drive not to delete is therefore equally strong for all these vowels.

However, it is informative to compare the set of non-deletion-undergoers. For any given input vowel, a non-undergoing candidate will violate different markedness

constraints, depending on whether it is occurs in final position of a prosodic word or elsewhere in a prosodic word. To make this more concrete, consider as an example the vowel /o/. This vowel reduces to the high back vowel in unstressed syllables. When the reduced vowel occurs in prosodic word final position, it will violate the markedness constraints *ŏ/u and *ŭ]$_{\omega}$. However, when it occurs elsewhere in the prosodic word, it violates only *ŏ/u. In prosodic word final position, deletion will avoid two markedness violations, while it will avoid only one elsewhere in the prosodic word. The drive to delete is therefore stronger final in the prosodic word than elsewhere, and a higher rate of deletion is predicted for this position. This is illustrated in the tableau in (59). Since the faithfulness violations incurred by the reduction candidate is the same irrespective of where in the prosodic word the vowel occurs, only the markedness violations are relevant here. In this tableau I include only markedness constraint actually violated by one of the candidates under consideration.

(59) **Non-deletion candidates from different positions in the prosodic word**

			*ŭ]$_{\omega}$	*ŏ/u
Elsewhere:	1	/o/ → [ŭ]		*
Prosodic word final:	2	/o#/ → [ŭ#]	*	*

Output of EVAL

Retention elsewhere /o/ → [ŭ]

Retention final in prosodic word: /o# → [ŭ#] *ŭ]$_{\omega}$

The discussion has been in terms of the input /o/. However, the same can be shown for the vowels /e, ɛ, ə, ɔ, u/. For every input vowel, non-deletion is more marked in prosodic word final position than elsewhere. For each of these vowels, the drive to

delete is therefore stronger in prosodic word final position, and we predict a general higher deletion rate in prosodic word final position.

This prediction also follows from the analysis of the deletion process presented above in §3 where ordinary generated comparison sets were discussed. Consider first the vowels that are realized as [ŭ]. In §3.2.1 I have shown how EVAL imposes for any one of the vowels /o, ɔ, u/ the following rank-ordering on the possible outputs when not in prosodic word final position: |ŭ ™ ∅|. Since [ŭ] is predicted to occur more than ∅ when not final in a prosodic word, and since these are the only two variants predicted as possible, it follows that deletion is predicted to occur less than 50% of the time. In §3.2.2, however, we have seen that EVAL imposes the opposite ordering on the retention and the deletion candidates when final in the prosodic word, i.e. |∅# ™ ŭ#|. Again, these are predicted to be the only two variants, and since the deletion candidate is the more frequent variant, deletion is predicted to occur more than 50% of time. When non-final in the prosodic word, deletion is predicted to happen less than half of the time; when final in the prosodic word, deletion is predicted to happen more than half time. This implies more deletion in prosodic word final position than elsewhere in the prosodic word. The same explanation can be given with regard to the deletion rate associated with [ə̆]. The fact the prediction based on generated comparison sets (§3) and the prediction based on non-generated comparison sets (the current section) are in agreement, is strong evidence in favor of the analysis.

4.3 Stress of the following syllable

A vowel in an unstressed syllable is more likely to delete if it is followed by an unstressed syllable (48%) than if it is followed by a stressed syllable (25%). In this

section I will argue that this follows form general principles of rhythmic well-formedness. There is a very strong universal tendency for rhythmic structure to be alternating – i.e. contiguous unstressed syllables are avoided, as are contiguous stressed syllables. In order to express this generalization, two constraints have been defined in OT, namely *LAPSE against contiguous unstressed syllables, and *CLASH against contiguous stressed syllables (Alber, 2002, Elenbaas and Kager, 1999, Kager, 2001, Kenstowicz and Green, 1995, Pater, 2000). In this section I will argue that it is these constraints that lead to higher deletion rates in pre-unstressed syllables. As in the preceding two sections, there are two non-generated comparison sets to consider, the undergoers and the non-undergoers. I will first discuss the non-undergoers.

4.3.1 The non-undergoers and *LAPSE

When an unstressed syllable is followed by an unstressed syllable, two contiguous unstressed syllables is the result. Such a form will violate the constraint *LAPSE against rhythmic lapses. However, when an unstressed syllable is followed by a stressed syllable, there is an alternating rhythmic structure, and therefore *LAPSE is not violated. In addition to general constraints against vowels in unstressed syllables, a candidate in which an unstressed vowel is followed by yet another unstressed syllable also violates *LAPSE. In such a form both *LAPSE and the markedness constraint against unstressed vowels will favor deletion, predicting higher deletion rates in such a context.

As an example, consider the vowel /e/ in two contexts: (i) Where both /e/ and the following vowel will be realized as unstressed, i.e. /e…v/ → [ə̆ … v̆]; and where /e/ will be realized as unstressed but the following vowel a stressed. /e…v/ → [ə̆…v́]. There exists a non-generated comparison set that contains these two mappings. The tableau in

(60) compares these two mappings. Since both mappings violate the same faithfulness constraints, the tableau includes only markedness constraints. I am also not considering the violations that the vowel in the following syllable might receive.[49]

(60) **Non-deletion in pre-stressed and pre-unstressed contexts**

			*P/ə	*LAPSE
Pre-unstressed:	2	/e…v/ → [ə̆ … v̆]	*	*
Pre-stressed:	1	/e…v/ → [ə̆…v́]	*	

Output of EVAL

Pre-stressed retention: /e…v/ → [ə̆…v́]

|

Pre-unstressed retention: /e…v/ → [ə̆ … v̆] *LAPSE

The comparison has been done for the vowel /e/ here, but the same comparison can be done for any of the vowels. For all vowels, non-deletion before an unstressed vowel will result in a more marked output than non-deletion before a stressed vowel. Deletion before an unstressed vowel can therefore avoid violation of two markedness constraints, *LAPSE and the markedness constraint against the unstressed vowel. However, deletion before a stressed vowel avoids violation of only one markedness constraint, namely against the unstressed vowel. The drive to delete is stronger before unstressed vowels, and the prediction is that this context should be associated with higher deletion rates. This is in agreement with the actual pattern observed in Silva's corpus.

4.3.2 The undergoers and *CLASH

Deleting an unstressed vowel violates MAX, irrespective of whether the vowel was followed by a stressed or an unstressed syllable. The faithfulness cost is therefore the

[49] About where *LAPSE should be ranked, see §4.3.3 below. For now I do not rank it relative to *P/ə.

same for both contexts under consideration here. However, deleting a vowel before a stressed vowel can result in bringing two stressed syllables into contact, therefore creating a *CLASH-violation. This will not always happen, but only when the deleting vowel is flanked by two stressed syllables, i.e. /σ́ v̆ σ́/ → [σ́ σ́].[50] When the deleting vowel is preceded by an unstressed vowel, then its deletion does not create a *CLASH-violation, i.e. /σ̆ v̆ σ́/ → [σ̆ σ́]. On the other hand, deletion of a vowel followed by an unstressed syllable can never cause a *CLASH violation, /σ́ v̆ σ̆/ → [σ́ σ̆] and /σ̆ v̆ σ̆/ → [σ̆ σ̆].[51] When the preceding syllable is stressed, then the markedness cost of deletion in a pre-stressed context is higher than the markedness cost of deletion in a pre-unstressed context. In addition to the faithfulness constraint MAX, *CLASH also militates against deletion in (some) pre-stressed contexts, with the result that deletion will apply less frequently before stressed syllables.

As an example, consider the input /ɔ/. There is a non-generated comparison set for this vowel that contains a mapping where it is deleted between two stressed syllables (/σ́ ɔ σ́/ → [σ́ σ́]), and a mapping where it is deleted between a stressed and an unstressed

---

[50] Since stress assignment in Portuguese is for the most part predictable, it is probably not correct to indicate stress on underlying representations. Therefore, whenever I indicate stress (or the lack of stress) on underlying form it should be interpreted as follows: This syllable will receive stress (or not) according to the rules of stress assignment of Portuguese.

[51] /σ̆ v̆ σ̆/ → [σ̆ σ̆]: The output form here does contain a *LAPSE violation. However, had deletion not applied, then the form would have had two violations of *LAPSE. In addition to avoiding the violation of the constraint against the unstressed vowel, deletion also avoids one *LAPSE-violation here.

/σ̆ v̆ σ́/ → [σ̆ σ́]: This is the corresponding mapping with a following stressed syllable. The output here does not violate *LAPSE. On the face of it, this seems to predict that deletion should apply more in pre-stressed position. However, had deletion not applied in this situation *LAPSE would have been violated once. Deletion therefore avoids violation of the constraint against the unstressed syllable, and one *LAPSE-violation.

When the syllable preceding the deleting vowel is unstressed, there is no difference between the two contexts in what is gained by deletion. The drive to delete is therefore equally strong when the preceding syllable is unstressed.

syllable (/σ́ ɔ σ̆/ → [σ́ σ̆]). The tableau in (61) compares the members of this non-generated comparison set.[52]

(62) **Comparing deletion candidates in pre-stressed and pre-unstressed contexts**

			MAX	*CLASH
Pre-stressed:	2	/σ́ ɔ σ́/ → [σ́ σ́]	*	*
Pre-unstressed:	1	/σ́ ɔ σ̆/ → [σ́ σ̆]	*	

Output of EVAL

Pre-unstressed deletion: /σ́ ɔ σ̆/ → [σ́ σ̆]

|

Pre-stressed deletion: /σ́ ɔ σ́/ → [σ́ σ́] *CLASH

This illustration has been in terms of the vowel /ɔ/. Since the constraints involved are independent of the input vowel, the same point can be made for any input vowel. When a vowel is preceded and followed by a stressed syllable, then deletion of that vowel will result in a *CLASH-violation. However, when it is preceded by a stressed syllable and followed by an unstressed syllable, then deletion of that vowel will not result in the creation of a *CLASH-violation. In these contexts, there are more constraints that militate against deletion before a stressed syllable. This therefore predicts higher deletion rates in pre-stressed position, which is in agreement with the patterns observed in Silva's corpus.

By taking into consideration the rhythmic well-formedness constraints *LAPSE and *CLASH, we can account for the fact that deletion is observed more frequently in pre-unstressed than in pre-stressed position. There are two reasons: (i) Non-deletion in a pre-unstressed position can result in a form that violates *LAPSE, while non-deletion in a pre-stressed position cannot. The drive to delete is therefore stronger in pre-unstressed

[52] About where *CLASH should be ranked, see §4.3.3 below.

position. (ii) Deletion in pre-stressed position can create a *CLASH-violation, while deletion in pre-unstressed position cannot. The drive not to delete is stronger in pre-stressed position than in pre-unstressed position.

4.3.3 *LAPSE, *CLASH and the deletion of individual vowels

In the previous two sections I have shown that *LAPSE and *CLASH can influence the deletion rate. In §3 above, I have offered an account of the variable deletion of the Portuguese vowels without taking into consideration the contribution of these two constraints. Addition of *LAPSE and *CLASH can cause problems for the earlier account. In §3 the ranking between MAX and the markedness constraints determines whether deletion or retention is preferred. If the markedness constraint violated by the retention candidate ranks above MAX, then deletion is preferred. If the markedness constraint violated by the retention candidate is ranked below MAX, then retention is preferred.

(63) **The structure of the argument from §3**

a. Markedness o MAX = Deletion preferred

||*v̌]$_{\omega}$ o MAX|| = more deletion than retention when final in prosodic word.

b. MAX o Markedness = Retention preferred

||MAXo*P/əo*ŏ/u|| = for [ə̆], [ŭ] more retention than deletion elsewhere.

If *CLASH and *LAPSE rank too high, they can override the effects of MAX and the constraints against unstressed vowels. They can do this in particular if they are allowed to rank above MAX. I illustrate this here with *LAPSE and an input /u/ followed by an unstressed syllable in a non-prosodic word final context. In this context, the retention candidate should be the more frequently observed output. However, under the ranking ||*LAPSE o MAX||, the opposite is predicted. This is shown in the tableau in (64).

(64) **The wrong predictions with ||*LAPSE o MAX||**

/u ɔ̆/	*LAPSE	MAX	*ɔ̆/u
2 ŭ ɔ̆	*		*
1 ∅ ɔ̆		*	

Output of EVAL

∅ ɔ̆ MAX

ŭ ɔ̆ *LAPSE

Cut-off

The illustration here was in terms of *LAPSE and /u/. However, similar examples could be constructed with the other vowels and with *CLASH. We therefore need these two constraints to be ranked below MAX.

4.4 Final summary

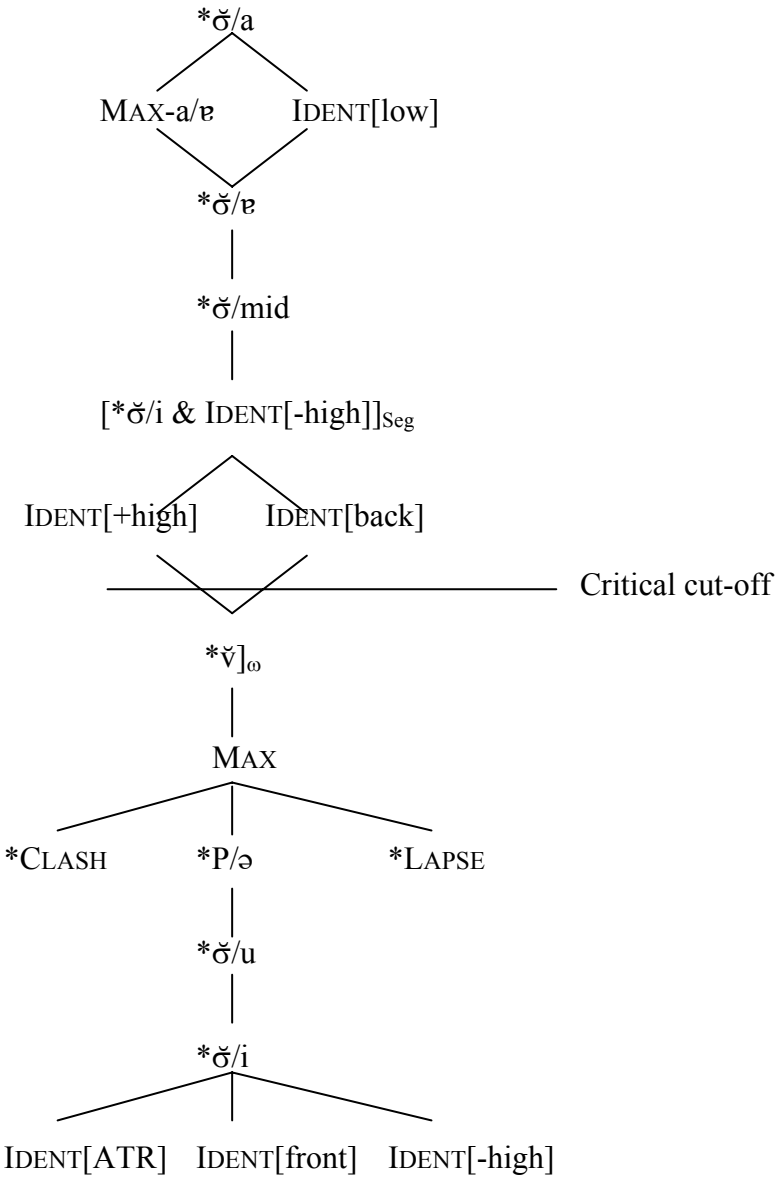
In the table in (65) I summarize all of the new rankings that have been argued for in §4. As before, I also give a graphic representation of Faialense the Portuguese constraint hierarchy. This is the final hierarchy for Faialense Portuguese.

(65) **Summary of new rankings**

Ranking	**Motivation**	**Where?**
*P/ə o *ɔ̆/u	[ə̆] deletes more than [ŭ]	§4.1.1 (55) (56)
MAX o*P/ə	In general [ə̆] is preferred over deletion	§4.1.1.1 (57)
MAX o {*CLASH, *LAPSE}	*CLASH, *LAPSE cannot override effect of MAX and markedness against unstressed vowels	§4.3.3

With these additional rankings, and by allowing EVAL to compare candidates that are not related to each by the same input (selected comparison sets), we can account for the following generalizations about vowel deletion in Faialense Portuguese: (i) some vowels are more prone to delete than others; (ii) vowels in prosodic word final position are more prone to delete than vowels elsewhere; and (iii) vowels before an unstressed syllable are more prone to delete than vowels before a stressed syllable.

(66) **Graphic representation of the Faialense Portuguese constraint hierarchy**



5. Summary: vowel reduction and deletion in Faialense Portuguese

In this chapter I have presented an account of vowel reduction and deletion in Faialense Portuguese within a rank-ordering model of EVAL. A rank-ordering model of EVAL differs in two respects from the classic OT conception of EVAL: (i) In the rank-ordering model, EVAL does more than to distinguish the optimal candidate from the losers. EVAL imposes a harmonic ordering on the complete candidate set. (ii) EVAL is allowed to compare candidates that are not related to each other via the same input. The rank-ordering model of EVAL can be used to account for variable phenomena in the following way:

(i) The language user can potentially access all candidates in the rank-ordering that EVAL imposes on the candidate set. However, the likelihood of a candidate being accessed depends on the position it occupies in this rank ordering. The higher position a candidate occupies in the rank-ordering, the more likely it is to be accessed. In a variable phenomenon the language user therefore accesses more than just the best candidate. However, the best candidate is most likely to be accessed and is therefore predicted to be to the preferred variant. The second best candidate is second most likely to be accessed, and is therefore the second most frequent variant, etc.

(ii) There is an absolute cut-off on the constraint hierarchy of every language. When given a choice a language user will not select as output a candidate that is disfavored by a constraint ranked higher than the critical cut-off. This places a strict limit on the number of variants that will be observed in a variable process, and it also accounts for categorical phenomena.

(iii) A variable process often applies with different frequency in different contexts. By allowing EVAL to compare candidates from different contexts (candidates that are not related to each other via a shared input), we can account for the relative frequency with which a variable process applies in different contexts.

Deletion of unstressed vowels applies variably in Faialense Portuguese. In this chapter I have shown how the variable deletion pattern in Faialense Portuguese can be accounted for within such a rank-ordering model of EVAL. In particular, I have shown how this model accounts for the following facts about vowel deletion in Faialense Portuguese:

(i) In non-prosodic word final position the unstressed vowels [ŭ, ə̆, ĭ] are retained more often than they are deleted. The grammar developed for Faialense Portuguese in this chapter rates the retention candidate best and the deletion candidate second best in non-prosodic word final position.

(ii) In prosodic word final position the unstressed vowels [ŭ, ə̆, ĭ] are deleted more often than they are retained. In this context the grammar rates the deletion candidate best and the retention candidate second best.

(iii) All other candidates for these vowels are eliminated by constraints ranked above the critical cut-off, and are therefore never selected as possible variants.

(iii) The unstressed vowel [ĕ] is categorically retained. All candidates for this output are disfavored by constraints above the cut-off. Only the single best candidate is therefore selected as output. The grammar rates [ĕ] as best.

(iv) [ə̆] deletes most frequently, then [ŭ], and then [ĭ]. Comparison between non-deletion candidates for these vowels show that [ə̆] is more marked than [ŭ] which is again more marked than [ĭ]. A higher deletion rate is predicted with more marked forms.

(v) Comparison between unstressed vowels in prosodic word final position and unstressed vowels from elsewhere in the prosodic word, show that the vowels in prosodic word final position are more marked. This explains why this position is associated with higher deletion rates.

(vi) Unstressed vowels followed by an unstressed syllable are more prone to delete than unstressed vowels followed by a stressed syllable. Comparison between candidates from these two contexts show that deletion in pre-unstressed position leads to more well-formed rhythmical structures, while deletion in pre-stressed syllables could create rhythmically more marked structures. This explains why pre-unstressed vowels are more prone to deletion.

CHAPTER 5

[t, d]-DELETION IN ENGLISH

In English, a coronal stop that appears as last member of a word-final consonant cluster is subject to variable deletion – i.e. a word such as *west* can be pronounced as either [wɛst] or [wɛs]. Over the past thirty five years, this phenomenon has been studied in more detail than probably any other variable phonological phenomenon. Final [t, d]-deletion has been studied in dialects as diverse as the following: African American English (AAE) in New York City (Labov et al., 1968), in Detroit (Wolfram, 1969), and in Washington (Fasold, 1972), Standard American English in New York and Philadelphia (Guy, 1980), Chicano English in Los Angeles (Santa Ana, 1991), Tejano English in San Antonio (Bayley, 1995), Jamaican English in Kingston (Patrick, 1991) and Trinidadian English (Kang, 1994), etc.[1] Two aspects that stand out from all these studies are (i) that this process is strongly grammatically conditioned, and (ii) that the grammatical factors that condition this process are the same from dialect to dialect. Because of these two facts [t, d]-deletion is particularly suited to a grammatical analysis. In this chapter I provide an analysis for this phenomenon within the rank-ordering model of EVAL.

The factors that influence the likelihood of application of [t, d]-deletion can be classified into three broad categories: the following context (is the [t, d] followed by a consonant, vowel or pause), the preceding context (the phonological features of the consonant preceding the [t, d]), the grammatical status of the [t, d] (is it part of the root or

---

[1] This phenomenon has also been studied in Dutch – see Schouten (1982, 1984) and Hinskens (1992, 1996). With a few exceptions the factors determining the likelihood of deletion in Dutch are virtually identical to those observed in English.

is it a suffix).[2] The contribution of each of these three factors can be summarized as follows: (i) *The following context*. [t, d] that is followed by a consonant is more likely to delete than [t, d] that is followed by either a vowel or a pause. Dialects differ from each other with regard to the influence of following vowels and pauses. In some dialects, a following vowel is associated with higher deletion rates than a following pause. In other dialects this situation is reversed – i.e. more deletion before a pause than a vowel. (ii) *Preceding context*. In general, the more similar the preceding segment is to [t, d], the more likely [t, d] is to delete. Similarity has been measured in terms of sonority (higher deletion rates after obstruents than sonorants), but also in terms of counting the number of features shared between [t, d] and the preceding consonant. (iii) *Grammatical category*. Generally speaking, [t, d] that is part of the root (in a monomorpheme like *west*) is subject to higher deletion rates than [t, d] that functions as a suffix (the past tense suffix in *locked*).

Of these factors, the first two can be classified as phonological and the third as morphological. Even though I acknowledge that morphology interacts with the process of [t, d]-deletion, I will discuss only the two phonological factors here. The rest of the chapter consists of a discussion within a rank-ordering model of EVAL of the following phonological context in §1, and the preceding phonological context in §2. There are at least three alternative accounts of variation in the OT literature. In section §3 I discuss these alternatives and show how they compare to the rank-ordering model of EVAL.

---

[2] In a summary statement of the phenomenon Labov (1989:89-90) actually identifies three additional factors, namely (i) whether the syllable to which the [t, d] belongs is stressed or unstressed, (ii) whether the cluster that the [t, d] belongs to consists of two or more consonants, and (iii) the voicing of the segments flanking the [t, d]. These three factors are less robust. Many studies that do report on these factors have found them not to contribute significantly to the likelihood of [t, d]-deletion. Several studies do not even report on these factors.

I assume familiarity with the rank-ordering model of EVAL in this chapter. For a general discussion of this and an illustration of how variation is accounted for in this model, refer to Chapter 1 and Chapter 3.

1. The following phonological context

One of the aspects that influence the rate of [t, d]-deletion is the nature of what follows the word-final [t, d]. The basic generalization can be stated as follows: (i) A [t, d] followed by a consonant is more likely to delete than a [t, d] followed by either a vowel or a pause. (ii) Dialects differ with respect to the relation between a following vowel and a following pause. In some dialects a following vowel is associated with higher deletion rates than a following pause, while in other dialects a following pause is associated with higher deletion rates than a following vowel. In this section I present an analysis of the effect of the following context on the [t, d]-deletion rate.

I will analyze the influence of the following context within the "licensing by cue" approach to phonological neutralization (Steriade, 1997). According to this approach, a sound is more likely to be neutralized in a context where it is more difficult to perceive the sound accurately. The assumption is that sounds are perceived/identified based on acoustic cues to their identity. However, not all cues are equally robust in all contexts. For instance, one of the cues for identifying the place of articulation of a consonant is the formant transitions from the consonant into a following vowel. This cue for place of articulation is therefore licensed in pre-vocalic position. However, if a consonant is not followed by a vowel, this cue for identifying the place of articulation of the consonant is not available as robustly. Place of articulation is therefore licensed more robustly in pre-

vocalic than, for instance, pre-consonantal position. Consequently, place of articulation is more likely to be neutralized in pre-consonantal than in pre-vocalic position.[3]

I will claim that stop consonants are less robustly licensed in pre-consonantal position than either pre-vocalic or pre-pausal position. This is the reason why [t, d] in pre-consonantal position deletes (is neutralized) more frequently than [t, d] in pre-pausal or pre-vocalic position. The robustness of licensing in pre-vocalic and pre-pausal position is subject to dialectal variation. In some dialects stop consonants are licensed more robustly pre-vocalically than pre-pausally, and *vice versa* in other dialects. This explains why some dialects delete [t, d] more before vowels and others more before pauses.

The rest of this section is structured as follows: In §1.1 I will present a selection of the data from the literature on the influence of the following context on [t, d]-deletion. These data will be analyzed within the rank-ordering model of EVAL in §1.2. Finally, in §1.3 I will consider alternative explanations for the influence of the following context. In particular, an analysis will be considered that relies on re-syllabification across word boundaries rather than on licensing by cue.

1.1 The data

The table in (1) contains a representative sample of the data on how [t, d]-deletion interacts with the following phonological context. Before discussing the pattern observed in these data, I will first give background on how the data were collected.

---

[3] This is indeed true. When a two consonant cluster occurs inter-vocalically, it is usually the first consonant that assimilates in place to the second – i.e. the place of the pre-vocalic consonant is preserved while the place of the pre-consonantal consonant is neutralized. See, for instance, place assimilation between the English negative prefix /in-/ and labial or velar initial roots: *i*[mp]*ratical* and *i*[ŋk]*onclusive*.

(1) **The influence of following context on [t, d]-deletion (in percentage)**[4]

		Pre-C	Pre-V	Pre-Pause
Chicano English (Los Angeles)	*n*	3,693	1,574	1,024
	% deleted	62	45	37
Tejano English (San Antonio)	*n*	1,738	974	564
	% deleted	62	25	46
AAE (Washington, DC)	*n*	143	202	37
	% deleted	76	29	73
Jamaican mesolect (Kingston)	*n*	1,252	793	252
	% deleted	85	63	71
Trinidadian English	*n*	22	43	16
	% deleted	81	21	31
Neu data	*n*	814	495	–
	% deleted	36	16	–

The data on the Chicano English are from Santa Ana (1991:76, 1996:66). These data are based on 45 speakers of Chicano English in the Barrio of Los Angeles that represent a balanced sample in terms of age and socio-economic status.

The data on Tejano English were collected by Bayley (1995:310). The data are from 32 speakers of Tejano English, all of whom live in the same housing project in San Antonio.

[4] Guy (1980) also reports on the English spoken by white Philadelphians and white New Yorkers. Unfortunately, he only reports the VARBUL factor values associated with the different contexts and not the actual deletion rates. Since it is not possible to determine the deletion rates based on the factor weights, I cannot use Guy's data. One thing that is clear from Guy's data, however, is that also in these two dialects pre-consonantal position is associated with higher deletion rates than pre-vocalic and pre-pausal position.

Labov *et al.* (1968:102) report on the English spoken by African Americans and Puerto Ricans in Harlem. Unfortunately, they lumped pre-consonantal and pre-pausal contexts together. They found high deletion rates before consonants and pauses (85% out of 1,929 tokens), and low deletion rates before vowels (34% out of 992 tokens).

The Washington, DC AAE data are from Fasold (1972:76). It is based on data collected from 51 speakers from a wide range of socio-economic classes and ages. Fasold considered only [t, d] that served as past tense markers – i.e. these data do not include deletion rates in monomorphemic words such as *west*.

The Jamaican data were collected by Patrick (1991:181) from 10 speakers of the Jamaican mesolect spoken in the Veeton suburb of Kingston. The speakers are representative of the social and educational classes of the community.

The data on Trinidadian English are from Kang (1994:157), and are based on the English of 13 middle class male speakers of standard Trinidadian English.[5]

The last set of data in data in (1) is from Neu (1980:45). It is based on the speech of 15 speakers from diverse backgrounds. Even so, Neu claims that these data reflect a sample from a homogeneous population. She performed several chi-square tests to test the null hypothesis that all the speakers showed the same deletion patterns (p. 41). The null hypothesis could never be rejected. Neu unfortunately did not report on pre-pausal context.

Now we can consider the patterns that are visible in these data. The data show that in all of these dialects of English, pre-consonantal context is associated with higher deletion rates than both pre-vocalic and pre-pausal context. It also shows that in some

---

[5] The very low number of tokens makes these results somewhat tentative. However, the general pattern agrees with the pattern observed in other dialects of English (more deletion in pre-consonantal than pre-vocalic or pre-pausal context). We can therefore tentatively accept these data. Kang also reports on the mesolectal and basilectal versions of Trinidadian Creole. However, these varieties of Trinidadian show very high deletion of final [t, d] across the board, something that Kang attributes to "strict syllable structure constraints … which rarely allow syllable-final consonant clusters" (p. 155). The combination of the high deletion rate and the very small number of tokens in Kang's corpus results in insignificant differentiation between the different contexts.

dialects [t, d] in pre-pausal context deletes more than [t, d] in pre-vocalic context (Tejano, Washington, DC AAE, Jamaican). In other dialects [t, d] in pre-vocalic context deletes more often than [t, d] in pre-pausal context (Chicano). The dialects can therefore be divided into two classes with deletion rates related as follows (in order of declining deletion rate): (i) pre-consonantal > pre-pausal > pre-vocalic, (ii) pre-consonantal > pre-vocalic > pre-pausal.

The data from table (1) are represented in a different format in (2). In this table the three contexts for each dialect are arranged according to the deletion rate associated with each context. Contexts with higher deletion rates occur to the left, and contexts with lower deletion rates occur to the right. A broken vertical line is drawn to indicate the 50%-mark. Contexts to the left of these lines are associated with deletion rates of above 50%, and contexts to the right of these lines show less than 50% deletion.

(2) **The influence of following context on [t, d]-deletion**

(Pre-C = pre-consonantal, Pre-V = pre-vocalic, Pre-P = pre-pausal.)

	More deletion ←			50% ↓	Less deletion →	
Chicano (Los Angeles)				Pre-C	Pre-V	Pre-P
Tejano (San Antonio), Trinidian				Pre-C	Pre-P	Pre-V
AAE(Washington, DC)			Pre-C	Pre-P	Pre-V	
Jamaican (Kingston)		Pre-C	Pre-P	Pre-V		
Neu data					Pre-C	Pre-V

In the rank-ordering model of EVAL we account for two aspects of variation: (i) *Inter-contextual variation*. For a specific input, which of the two variants (deletion or retention) is the more frequently observed variant? For the English dialects discussed here, this information can be read off from the table in (2) as follows: If a context appears

to the left of the 50% mark, then deletion is preferred is the more frequent variant. If it appears to the right of the 50% mark, then retention is the more frequent variant. (ii) *Intra-contextual variation*. Here we compare inputs from different contexts and ask which context has the higher deletion rate. This information can also be read off table (2). If for some dialect $context_1$ appears to the left of $context_2$, then $context_1$ has a higher deletion rate than $context_2$.

1.2 The analysis

In this section I will provide an analysis for the data in table (2). The section will start out in §1.2.1 with a discussion of the constraints involved. In §1.2.2 I will then show how the constraints can be ranked to account for each of the patterns exemplified in (2). Section §1.2.3 will consider the factorial typology that is predicted by the constraints that I propose in §1.2.1 – i.e. in addition to the dialects in table (2), which other dialects are predicted to be possible? Finally, in §1.2.4 I will discuss two outstanding questions: (i) does only [t, d] delete, and (ii) does word-final [t, d] also delete post-vocalically?

1.2.1 The constraints

Since the unfaithful mapping that we are dealing with here is one of deletion, the faithfulness constraint involved in explaining this pattern is the anti-deletion constraint MAX. In order to explain why this constraint is sometimes violated, we need markedness constraints that would be violated by the faithful non-deletion candidates. I will argue that the relevant markedness constraints are contextual licensing constraints (Steriade, 1997).

This type of constraint was first formulated to explain the contextual distribution of phonological contrasts. The idea is that a contrast is preserved (licensed) more easily in contexts where the cues for its perception are more salient than in contexts where these cues are less salient. For instance, one of the most salient cues for the voicing distinction in stop consonants is voice onset time (VOT) (Lisker and Abramson, 1964, Lisker, 1986). VOT is the time that elapses between the release of consonantal closure and the onset of voicing. VOT is more robustly licensed before a sonorant segment such as a vowel than before a word boundary. The result is that voicing is more easily neutralized before a word-boundary than before a vowel. Steriade (1997) captures this generalization by formulating a markedness constraint against voicing in each of these two contexts, i.e. *voice/ __ [+sonorant] (violated *inter alia* by a voiced stop in pre-vocalic position) and *voice/ __ # (violated by a voiced stop before a word-boundary). Steriade further argues that the constraint against voicing in the less robustly licensing context universally outranks the constraint against voicing in the more robustly licensing context, i.e. ||*voice/ __ # o *voice/ __ [+sonorant]||. A word-final voiced stop then violates a higher ranking markedness constraint than a pre-vocalic voiced stop. This implies that a word-final voiced stop is more likely to devoice than a pre-vocalic voiced stop.

In this section I am distinguishing three different word-final contexts that interact with the likelihood of [t, d]-deletion, namely pre-consonantal, pre-vocalic and pre-pausal. I propose a constraint against realizing a [t, d] in each of these three contexts. The three constraints are stated in (3a). Furthermore, based on the fact that pre-consonantal context is associated with higher deletion rates in all of the dialects studied, I claim that the

constraint against [t, d] in this context outranks the other two markedness constraints. This is shown in (3b). A motivation for these constraints and this ranking follows below.

(3) a. **Markedness constraints**

*Ct#C: A word-final [t, d] is not allowed if it is both preceded and followed by a consonant.

*Ct##: A word-final [t, d] is not allowed if it is preceded by another consonant and followed by a pause.

*Ct#V: A word-final [t, d] is not allowed if it is preceded by another consonant and followed by a vowel.

b. **Ranking**

||*Ct#C o {*Ct##, *Ct#V}||

In order to motivate the existence of the constraints in (3a) we need to determine what the cues are for identifying a [t, d], and then we need to show that these cues are differently realized in these three contexts. In order to motivate the ranking in (3b), we need to show that the cues for identifying [t, d] are less robustly realized in pre-consonantal context than in pre-pausal or pre-vocalic context.

There are two aspects of the identity of a [t, d] that need to be conveyed in order to distinguish [t, d] from other consonants: (i) its manner of articulation (to distinguish it from continuants and sonorants), and (ii) its place of articulation (to distinguish it from stops of other places of articulation). Since the environment preceding the [t, d] is the same in all three contexts under consideration, I will focus only on the following

environment – i.e. what are the cues for place and manner of articulation following a consonant?

In the literature two aspects are identified that can cue the place and manner of articulation of a consonant, namely the release of the consonant and the formant transitions from the consonant into a following sonorant. I will first discuss the cues available in consonant releases, and then the cues available in formant transitions.

Consonant releases. With regard to consonant releases, Stevens and Keyser argue that the most important distinctions in both place and manner of articulation can be cued successfully by how spectral energy distributions change at consonant releases (Stevens and Keyser, 1989:87). The distinction between [-continuant] and [+continuant] consonants is cued by the fact that non-continuants are characterized by an abrupt increase in amplitude over a range of frequencies at the release of the consonant. This is a result of the fact that the energy is absent (at most or all frequencies) during the closed phase of a non-continuant. In a continuant consonant energy is present over a wide range of frequencies during the complete consonantal pronunciation. There is therefore no abrupt rise in energy at the completion point of a continuant. With regard to place of articulation, Stevens and Keyser only comment on the distinction between coronal and non-coronal consonants. They argue that coronal consonants are cued by a greater increase in spectrum amplitude at high frequencies than at low frequencies at the termination point of the consonant. For non-coronals, spectrum amplitude is more likely to increase in the lower frequency ranges.

At least for place of articulation there is more evidence that the consonantal release carries enough information to cue the different places of articulation. See for

instance Lahiri *et al.* (1984) who show that it is possible to correctly distinguish between labial and coronal voiceless stops based on the shape of the spectral energy distribution from the consonantal release to the onset of voicing. See also Malécot (1958) who shows that word-final stop consonants are more accurately identified when they are released than when they are not. All of this serves as evidence that consonant releases do contain cues that can be used to identify [t, d].

Formant transitions. There is also ample evidence that formant transitions from a consonant into a following sonorant can cue both place and manner distinctions. For instance, Stevens and Blumstein (1978:1363) found that synthetic stops that were cued by only formant transitions were identified 72% more accurately than synthetic stops that were cued by only bursts (i.e. releases). See also Kewley-Port (1983) and Kewley-Port *et al.* (1983) (and their references) for evidence that such "time-varying" cues can be used to identify place of articulation.

There is a large body of literature on "locus equations" (e.g. Celdran and Villalba, 1995, Eek and Meister, 1995, Fowler, 1994, Fruchter and Sussman, 1997, Nearey and Shammass, 1987, Sussman *et al*., 1991, Sussman and Shore, 1996). A locus equation is the equation for a straight line that connects the second formant (*F2*) height at vowel onset with *F2*-height at the vowel midpoint. These equations are remarkably constant within consonants with the same place of articulation, so that it is possible to define a single equation that characterizes each place of articulation. On the other hand, the locus equations for consonants that differ in place of articulation are very different. Locus equations can therefore be used to classify consonants successfully in terms of their place

of articulation. This shows that formant transitions between a consonant and a following sonorant contain robust cues for identifying the place of articulation of the consonant.

What about manner of articulation? We have evidence that formant transitions can cue the distinction at least between stops and glides. Diehl and Walsh (Diehl and Walsh, 1989, Walsh and Diehl, 1991), for instance, have shown that the duration of the formant transitions can serve as a successful cue for the distinction between stops and glides. Longer formant transitions cue a glide percept, and shorter formant transitions cue a stop percept.

We therefore know that both the release of a [t, d] and the formant transitions from a [t, d] into a following segment contain cues for its identification. How are these cues realized in the three word-final contexts that interact with [t, d]-deletion? *Pre-vocalic.* In pre-vocalic position, both consonantal releases and formant transitions can be realized – i.e. both cues are potentially present in this context. However, realization of these cues requires that [t, d] be released into and transition into a vowel *across* a word boundary.[6] *Pre-pausal.* With a pre-pausal [t, d] the possibility of transitioning into a following vowel does not exist – since there is no vowel following the [t, d]. For a [t, d] in this context, the cues contained in the formant transitions are therefore not available at all. However, pre-pausal stops can be released.[7] The cues contained in the consonantal

---

[6] In this respect it differs from a consonant that precedes a vowel that is part of the same word – such a consonant can be released into and transition into the following vowel without crossing a word boundary. We can therefore expect that both release and formant transitional cues will be less robust for a word-final consonant followed by a vowel than for a consonant followed by a vowel that is part of the same word. This is borne out by patterns observed in neutralization processes – consonants in onset position (i.e. preceding a vowel that is part of the same word) is much less likely to undergo a neutralization process than a consonant in word-final position.

[7] See Holmes (1995:443) who claims that pre-pausal word-final /t/ is often aspirated in New Zealand English. Especially in aspirated stops, the release cues will be strongly present. Although I am not

release are therefore potentially available for [t, d] in pre-pausal position, and these release cues can be realized without crossing a word-boundary. *Pre-consonantal.* A pre-consonantal stop is practically never released,[8] so that the cues contained in the consonantal release are not generally available for pre-consonantal [t, d]. A consonant is also much less likely to have formant transitions into a following consonant than into a following vowel. In general, a consonant will only show formant transitions into a following sonorant consonant, and even then the transitions are less robust than transitions into a vowel. Weak formant transitional cues are therefore potentially available for a [t, d] followed by a sonorant consonant. As with the pre-vocalic context, formant transitions into a following sonorant would also require that the [t, d] transitions cross a word boundary. The table in (4) summarizes discussion.

The comparison in table (4) shows that pre-consonantal context is least likely to contain the cues necessary to identify [t, d]. Only transitional cues are potentially present. And even if they are present, they are present only weakly and only before a small subset of the consonants. This context is therefore the weakest in licensing the presence of [t, d].

---

aware of data showing that some American dialects aspirate pre-pausal word-final [t]'s, New Zealand English shows that it is at least possible.

[8] Browman and Goldstein (1990:363-366) analyze two utterances of the sequence "perfect memory". In the first utterance the words were pronounced with a pause between them. In the second utterance they were pronounced as part of a sentence – i.e. with no intervening pause. They say of the second utterance: "the final /t/ in 'perfect' is deleted in the traditional sense – careful listening reveals no evidence of the /t/" (p. 365). Browman and Goldstein tracked the movement of the tongue tip with X-ray. A movement of the tongue tip towards the alveolar ridge was interpreted as evidence that the coronal articulation was indeed performed. They found evidence for the coronal articulation in both articulations, and in particular they found that the articulation was of roughly equal magnitude in both utterances. The perceptual absence of the /t/ is therefore not due the coronal articulation not being made. The difference between the utterances is located in the release. In the first utterance there is very clear evidence of a release in the waveform, but about the second utterance they say that "no release can be seen in the waveform" (p. 365). The reason for this is that the articulation of the /m/ from "memory" partially overlaps with the articulation of the /t/ from "perfect". When the coronal closure is released the labial closure is therefore already in place. The result is that the coronal release has no aerodynamic/acoustic consequences. This lends evidence to the claim that the acoustic cues contained in the consonantal release are not generally available in pre-consonantal position.

This can be captured by ranking the constraint against [t, d] in this context higher than the constraint against [t, d] in the other contexts – as is done in (3b) above. No clear ordering can be established between pre-pausal and pre-vocalic context. In pre-vocalic context both release and transitional cues are available, but they require that the cues be realized across a word-boundary. In pre-pausal context only the release cues are available, but these cues can be realized without crossing a word-boundary. It can therefore be expected that there will be more freedom in how likely these two contexts are to sponsor a [t, d]. This can be captured by not imposing a universal ranking between the constraints against [t, d] in these contexts – as is done in (3b).

(4) **The presence of cues in different contexts**

__ # V		**__ ##**		**__ # C**	
Release	**Transition**	**Release**	**Transition**	**Release**	**Transition**
Yes Cross #	Yes Cross #	Yes	No	No	Weakly Pre-sonorant Cross #

1.2.2 Accounting for the observed patterns

There are two aspects of the variation pattern that we need to account for. (i) *Intra-contextual variation*. Consider the Chicano English data. In this dialect pre-consonantal context has a deletion rate of more than 50%. For a [t, d] input in which the final [t, d] occurs before a consonant, the deletion candidate therefore has to be rated as more well-formed than the retention candidate. The opposite is true in pre-vocalic and pre-pausal

contexts. Here less than 50% of [t, d] deletes, so that the retention candidate has to be rated better than the deletion candidate. This is the first aspect of the variation that we have to account for – for a specific input, which of the variants is the more frequently observed variant. Since dialects can differ in this respect, we have to consider each dialect individually. (ii) *Inter-contextual variation*. Consider again the Chicano dialect. Although [t, d] deletes less than 50% in both pre-vocalic and pre-pausal position, it deletes more in pre-vocalic than in pre-pausal position. We also have to account for this difference in deletion rates between the different contexts. In the rest of this section I first discuss the intra-contextual variation (§1.2.2.1) and then the inter-contextual variation (§1.2.2.2).

1.2.2.1 Intra-contextual variation

In the phenomenon of final [t, d]-deletion, two variants are observed – the retention candidate in which the [t, d] is preserved in pronunciation, and the deletion candidate (I will use ∅ to stand for this candidate). In contexts where [t, d] is observed more frequently than ∅, [t, d] has to be the more accessible candidate. EVAL must therefore rate [t, d] better than ∅ in that context, i.e. |t/d ™ ∅|. In contexts where ∅ is the more frequent variant, EVAL has to impose the opposite rank-ordering on these two candidates, i.e. |∅ ™ t/d|.

In all of the dialects discussed in §1.1 above variation is observed in all three contexts. This means that in all three contexts neither the deletion candidate ∅ nor the retention candidate [t, d] can be disfavored by any constraint ranked higher than the

critical cut-off. In all these dialects the cut-off is therefore located above MAX and the three markedness constraints from (3a).

Tableau (5) shows the violation profiles of a retention candidate and a deletion candidate in each of the three contexts. The only ranking that is assumed in this tableau is that *Ct#C outranks *Ct## and *Ct#V – see (3b) above. In particular, I make no claim yet about the ranking of MAX in this tableau – I indicate my agnosticism about the ranking of MAX by separating it from the other constraints by a squiggly line. This tableau does not yet represent the grammar of any specific dialect. It is presented here only to facilitate the discussion that follows. In this and all further tableau /Ct#C/ stands for an input where word-final [t, d] is followed by a consonant, /Ct#V/ where it is followed by a vowel, and /Ct##/ where it is followed by a pause.

(5) **Violation profiles of deletion and retention candidates in each of the contexts**

		*Ct#C	*Ct#V	*Ct##	MAX
/Ct#C/	t	*			
	∅				*
/Ct#V/	t		*		
	∅				*
/Ct##/	t			*	
	∅				*

Consider the pre-consonantal context. Suppose that we are dealing with a dialect in which the retention candidate were the preferred variant in this context – i.e. EVAL would have to impose the ordering |t ™ ∅| on the candidates in this context. In order to achieve this, the highest ranked constraint that distinguishes between the two candidates has to be a constraint that favors the retention candidate over the deletion candidate: MAX

has to outrank *Ct#C, or ||MAX o *Ct#C||. Now suppose that we were dealing with a different dialect in which the deletion candidate was the preferred variant in pre-consonantal context. Everything is now simply turned around. EVAL has to impose the ordering |∅ ™ t| on the candidate set, and in order to achieve that we need the ranking ||*Ct#C o MAX||. In fact, this gives a general heuristic for ranking MAX and the markedness constraints. Since I will use this heuristic over and over again in the discussion below, I state it explicitly in (6).

(6) **Heuristic for ranking MAX and the contextual licensing constraints**

Let *x* stand for some context, and CON-*x* for the markedness constraint against [t, d] in that context.

a. **|t ™ ∅|**

If in context *x* the retention candidate is the preferred variant, then:

||MAX o CON-*x*||.

b. **|∅ ™ t|**

If in context *x* the deletion candidate is the preferred variant, then:

||CON-*x* o MAX||.

Now we can consider the dialects separately. Let us start with Chicano, Tejano and Trinidadian English. In all three of these dialects the retention candidate is the preferred variant in pre-vocalic and pre-pausal contexts. However, in pre-consonantal context, the deletion candidate is observed more frequently than the retention candidate. In pre-consonantal position clause (6b) therefore applies to these dialects – i.e. we need the ranking ||*Ct#C o MAX||. However, in pre-pausal and pre-vocalic contexts, clause (6a)

applies – i.e. we need the ranking ||MAX o {*Ct#V, *Ct##}||. We still need to determine where to the critical cut-off should be located. Since variation between the deletion and retention candidates is observed in all three contexts, it means that neither of these candidates can be disfavored by a constraint ranked higher than the critical cut-off. This implies that MAX and all three markedness constraints rank lower than the cut-off.

(7) **Chicano, Tejano and Trinidadian English**

			DEP	*Ct#C	MAX	*Ct#V	*Ct##
/Ct#C/	2	t		*			
	1	∅			*		
	3	Vt	*!				
/Ct#V/	1	t				*	
	2	∅			*		
	3	Vt	*!				
/Ct##/	1	t					*
	2	∅			*		
	3	Vt	*!				

Output of EVAL

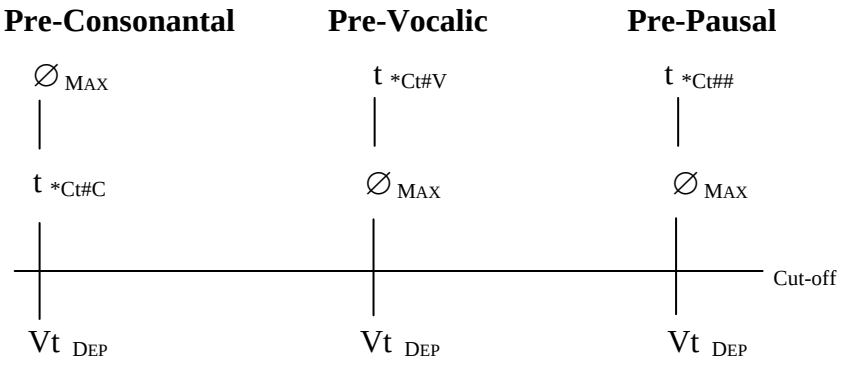


Tableau (7) represents a partial grammar of Chicano, Tejano and Trinidadian English. In addition to the deletion and the retention candidate, I also include a candidate that avoids the markedness violation by epenthesis. This candidate serves as an example

of a candidate that is never observed as a variant for any of the inputs considered here. The epenthetic candidate will also occupy a slot in the rank-ordering that EVAL imposes on the candidate set. To explain why the language user never accesses this candidate as a variant we have to call on the critical cut-off. The epenthetic candidate has to be disfavored by a constraint ranked higher than the critical cut-off. I therefore rank the anti-epenthesis constraint DEP higher than the critical cut-off. The epenthetic candidate stands in here as representative of all candidates never observed as variants – all such candidates will be disfavored by at least one constraint ranked higher than the cut-off. For the typographical conventions used in this tableau, see Chapter 1 §2.2.1 and §2.2.3

In pre-consonantal context EVAL imposes the rank-ordering |∅ ™ t ™ Vt| on the candidate set. Of these three candidates, the epenthetic candidate is disfavored by DEP which is ranked higher than the cut-off. Since there are candidates available that are not disfavored by any constraints ranked higher than the cut-off, this epenthetic candidate will never be selected as output. Of the two candidates that are possible outputs, the deletion candidate appears higher on the rank-ordering. It is therefore the more accessible of the two, and it is predicted to be the more frequently selected variant. In pre-vocalic and pre-pausal position, EVAL imposes the rank-ordering |t ™ ∅ ™ Vt| on the candidate set. In these contexts the epenthetic candidate is eliminated as output in the same manner as in the pre-consonantal context. Of the two possible outputs the retention candidate is here rated better and therefore more accessible. For these contexts, we expect more retention than deletion.

The other dialects are easily accounted for in a similar fashion. I will not discuss these other dialects in as much detail as these first three, since it is a straightforward

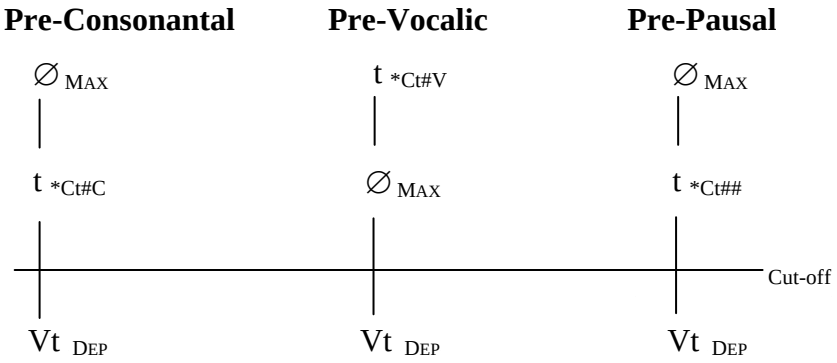
matter to arrive at the correct ranking for each dialect using the heuristic stated in (6). In the rest of this section I therefore only give the tableaux for the other dialects, with minimal discussion of each dialect.

In the AAE of Washington, DC the retention candidate is the preferred variant in pre-vocalic context. However, in both pre-consonantal and pre-pausal context, deletion is preferred over retention. The tableau for this dialect is given in (8).

(8) **Washington, DC AAE**

			DEP	*Ct#C	*Ct##	MAX	*Ct#V
/Ct#C/	2	t		*			
	1	∅				*	
	3	Vt	*!				
/Ct#V/	1	t					*
	2	∅				*	
	3	Vt	*!				
/Ct##/	2	t			*		
	1	∅				*	
	3	Vt	*!				

Output of EVAL

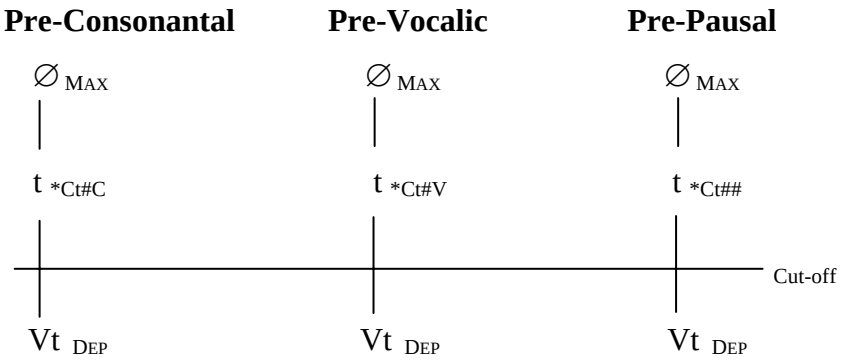


In Jamaican English deletion is preferred over retention in all three contexts. The tableau for this dialect is given in (9).

(9) **Jamaican English**

			DEP	*Ct#C	*Ct##	*Ct#V	MAX
/Ct#C/	2	t		*			
	1	∅					*
	3	Vt	*!				
/Ct#V/	2	t				*	
	1	∅					*
	3	Vt	*!				
/Ct##/	2	t			*		
	1	∅					*
	3	Vt	*!				

Output of EVAL



In the data reported by Neu (1980) retention is preferred over deletion in both pre-consonantal and pre-vocalic position. Since Neu did not report the pre-pausal deletion rates, I am not considering an input from this context. The tableau in (10) represents the grammar represented by these data.

This section has shown how intra-contextual variation can be accounted for in the rank-ordering model of EVAL. In each context, either the deletion or the retention candidate is the more frequent variant. For a context in which the deletion candidate is more frequent, the markedness constraint against [t, d] in that context outranks MAX, i.e.

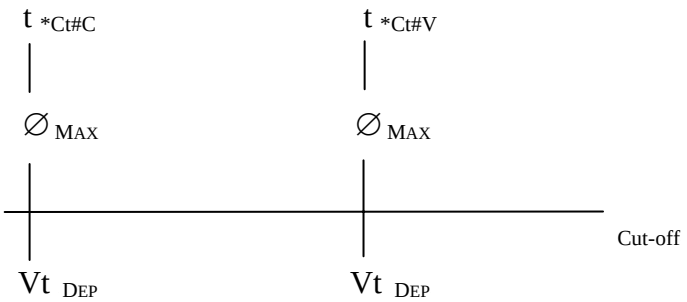
||Markedness o MAX||. With this ranking EVAL imposes the rank-ordering |∅ ™ t/d| on the candidate set. In a context where retention is more frequent, MAX outranks the markedness constraint, i.e. ||MAX o Markedness||. This results in EVAL imposing the opposite rank-ordering on the candidate set for that context: |t/d ™ ∅|.

(10) **The Neu data**

			DEP	MAX	*Ct#C	*Ct##	*Ct#V
/Ct#C/	1	t			*		
	2	∅		*			
	3	Vt	*!				
/Ct#V/	1	t					*
	2	∅		*			
	3	Vt	*!				

Output of EVAL

Pre-Consonantal **Pre-Vocalic**



But this still accounts for only one aspect of the variation pattern. Consider Chicano English as an example. Both in pre-vocalic and in pre-pausal position the retention candidate is the preferred variant. The ranking ||MAX o {*Ct#V, *Ct##}|| accounts successfully for this. However, although both of these contexts are associated with retention rates of more than 50%, the contexts do differ. Pre-vocalic

context has a higher deletion rate than pre-pausal context. This inter-contextual variation still has to be explained. This is the topic of the next section.

1.2.2.2 Inter-contextual variation

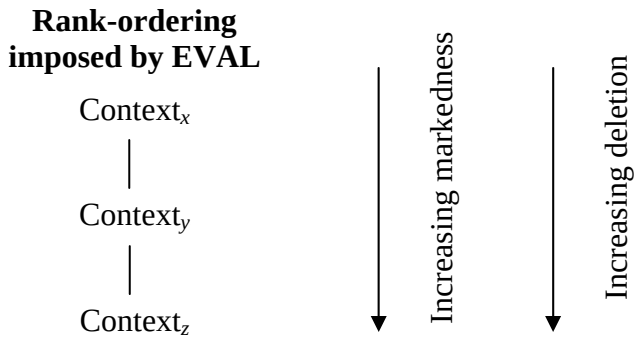
In order to account for inter-contextual variation we have to rely on the ability of EVAL to evaluate non-generated comparison sets (see Chapter 1 §1.2 and §2.2.2). One non-generated comparison set is of particular relevance in accounting for the inter-contextual variation in [t, d]-deletion. This is the set that contains the faithful candidates from each of the three different contexts, i.e. {/Ct#C/ → [t], /Ct#V/ → [t], /Ct##/ → [t]}.[9] EVAL will evaluate these three candidates in exactly the same way as it does ordinary generated comparison sets, and it will also impose a harmonic rank-ordering on these three candidates. Since these are the faithful candidates, they will only differ in terms of markedness violations. Consequently, the lower a candidate appears in the rank-ordering, the more marked it is.

The force that drives unfaithfulness is markedness. The more marked some faithful candidate, the stronger the drive to be unfaithful to that candidate. Therefore, the lower a candidate appears in the rank-ordering, the stronger the drive to delete, and the higher the deletion rate is expected to be in the context represented by that candidate. This discussion is represented graphically in (11).

---

[9] Another kind of non-generated comparison set can also be relevant, namely the set that contains the unfaithful (deletion) candidates from each of the three contexts: {/Ct#C/ → ∅, /Ct#V/ → ∅, /Ct##/ → ∅}. Of the constraints that we are considering here, all three of these candidates violate only MAX. EVAL can therefore not distinguish between them. In this particular instance, consideration of this non-generated comparison set is not informative.

(11) **Comparison between faithful candidates from different contexts**



In all dialects of English pre-consonantal position is associated with the highest deletion rate. This means that the drive to delete must be stronger in pre-consonantal context than in either pre-vocalic or pre-pausal context. Put in terms of a comparison between the faithful candidates: the faithful candidate from the pre-consonantal context must be more marked than the faithful candidate from both of the other two contexts. This can be achieved by ranking the markedness constraint against [t, d] in pre-consonantal position higher the constraints against [t, d] in the other two contexts, i.e. ||*Ct#C o {*Ct#V, *Ct##}||. This is indeed also the ranking that was argued for (3b) above based on the robustness of the cues for correctly identifying [t, d]. The cues for identifying [t, d] are least robust in pre-consonantal position, and therefore the constraint against [t, d] in this position was ranked the highest.

Dialects diverge in terms of deletion rates before vowels and before pauses. Although all dialects have lower deletion rates in these contexts than before consonants, some dialects delete more before a vowel than a pause (Chicano), and others delete more before a pause than before a vowel (Tejano, Washington, DC AAE, Trinidadian, Jamaican). This difference can be explained by ranking the constraints against [t, d] in pre-pausal and pre-vocalic position differently. In those dialects where pre-

vocalic position is associated with higher deletion rates, we need the faithful candidate from a pre-vocalic context to be more marked than the faithful candidate from a pre-pausal context. This can be achieved by the ranking ||*Ct#V o *Ct##||. On the other hand, in dialects where pre-pausal position is associated with more deletion, the faithful candidate from the pre-pausal context has to be more marked than the faithful candidate from the pre-vocalic context. This can be achieved by the ranking ||*Ct## o *Ct#V||. Referring back to (3b) will show that I did not claim a fixed ranking to exist between these two constraints. It does not seem to be the case that the cues for identifying [t, d] are inherently more robust in one of these contexts than the other.[10]

There are two kinds of dialects, and therefore two rankings between markedness constraints. These two kinds of dialects and the rankings associated with each are summarized in (12).

(12) **Different deletion rates in different contexts following [t, d]**

a. **Type A**

Deletion rates:	Pre-C > Pre-V > Pre-Pause
Dialects:	Chicano English
Ranking:	||*Ct#C o *Ct#V o *Ct##||

[10] It is not clear what determines which ranking is chosen in a specific dialect. It is possible that this is an arbitrary choice that has to be stipulated for every language. It is also possible that it can be related to finer details of phonetic implementation. Some dialects of English may more readily release stop consonants in pre-pausal position than other dialects. If this is true, then these pre-pausal release dialects will be dialects in which the cues in pre-pausal position are particularly robust. These could be the dialects in which in pre-pausal position is associated with lower deletion rates than pre-vocalic position. However, no data is available on whether pre-pausal stops are released or not in the dialects of English under investigation here. This therefore remains speculation for the time being. See Guy (1994:143) for similar speculation.

((12) continued)

b. **Type B**

Deletion rates: Pre-C > Pre-Pause > Pre-V

Dialects: Tejano, Washington, DC AAE, Jamaican, Trinidadian

Ranking: ||*Ct#C o *Ct## o *Ct#V||

As an illustration, I will discuss one dialect of each kind in more detail. I start with Chicano English as and example of dialect type A. In (7) above I have argued for the following partial ranking for this dialect: ||DEP o cut-off o *Ct#C o MAX o {*Ct#V, *Ct##}||. All that was missing to make this a complete ranking, is a ranking between the licensing constraints for pre-vocalic and pre-pausal position. Since pre-vocalic position is associated with higher deletion rates, we know that the constraint for this context has to outrank the constraint for the pre-pausal context, i.e. ||*Ct#V o *Ct##||. Tableau (13) considers the non-generated comparison set with the faithful candidates from the three contexts with this ranking added.

(13) **Chicano English: Comparing the faithful candidates**

		DEP	*Ct#C	MAX	*Ct#V	*Ct##
3	/Ct#C/ → [t]		*			
2	/Ct#V/ → [t]				*	
1	/Ct##/ → [t]					*

Output of EVAL

Pre-pausal: /Ct##/ → [t] $_{*Ct\#\#}$

Pre-vocalic: /Ct#V/ → [t] $_{*Ct\#V}$

Pre-consonantal: /Ct#C/ → [t] $_{*Ct\#C}$

This comparison shows that the faithful candidate in pre-pausal context is the least marked. Changing the input in this context (by deletion) will lead to the smallest decrease in markedness. The drive to be unfaithful is therefore the weakest in this context, and we are predicting the lowest deletion rate in this context. The faithful candidate from the pre-consonantal context is most marked, so that deleting the [t, d] from an input in this context will lead to the largest decrease in markedness. The drive to delete is strongest in this context, and we are predicting that this context will be associated with the highest deletion rates.

Consider Jamaican English as an example of a type B dialect. In this dialect the different contexts are related as follows in terms of deletion: Pre-C > Pre-Pause > Pre-V. In (9) I argued for the following hierarchy for this dialect: ||DEP o Cut-off o *Ct#C o {*Ct##, *Ct#V} o MAX ||. All we need to do is add the ranking between the constraints for pre-pausal and pre-vocalic contexts. Tableau (14) considers the non-generated comparison set with the faithful candidates from the three contexts for this dialect.

(14) **Jamaican English: Comparing the faithful candidates**

		DEP	*Ct#C	*Ct##	*Ct#V	MAX
3	/Ct#C/ → [t]		*			
1	/Ct#V/ → [t]				*	
2	/Ct##/ → [t]			*		

Output of EVAL

Pre-vocalic: /Ct#V/ → [t] $_{*Ct\#V}$

|

Pre-pausal: /Ct##/ → [t] $_{*Ct\#\#}$

|

Pre-consonantal: /Ct#C/ → [t] $_{*Ct\#C}$

The faithful candidate in pre-vocalic context is the least marked, so that deletion in this context will lead to the smallest decrease in markedness. We are expecting the lowest deletion rate in this context. As in Chicano English, the faithful candidate from pre-consonantal context is the most marked. Deletion in this context will buy the largest decrease in markedness, so that this context is expected to have the highest deletion rate.

By allowing EVAL to compare non-generated comparison sets, we can also account for inter-contextual variation. The basic idea is that unfaithfulness to the input is motivated only in order to decrease in terms of markedness. The more marked the faithful candidate is, the more can be gained by being unfaithful to the input. The highest deletion rates are therefore expected in contexts where the faithful candidate is most marked, and the lowest deletion rates in contexts where the faithful candidate is least marked.

1.2.3 Factorial typology – what are the possible dialects?

One of the claims of OT is that every possible ranking between the constraints represent a possible grammar, and therefore a possible language (or dialect in the current context). In the discussion above I have used four constraints to account for the [t, d]-deletion patterns observed in different dialects of English – MAX and three markedness constraints. I have also argued that the constraint against [t, d] in pre-consonantal position universally outranks the constraints against [t, d] in pre-vocalic and pre-pausal position – see (3b). This implies that there are two rankings possible between these markedness constraints. These two rankings are represented in (15).

(15) **Rankings possible between the markedness constraints**

a. ||*Ct#C o *Ct#V o *Ct##||

b. ||*Ct#C o *Ct## o *Ct#V||

In each of these two rankings there are four positions into which MAX can rank, so that each of these rankings represent four rankings with MAX included. I list the four rankings that (15a) stands for as an example in (16).

(16) **Including MAX in (15a)**

a. ||MAX o *Ct#C o *Ct#V o *Ct##||

b. ||*Ct#C o MAX o *Ct#V o *Ct##||

c. ||*Ct#C o *Ct#V o MAX o *Ct##||

d. ||*Ct#C o *Ct#V o *Ct## o MAX||

We also have to take into account the critical cut-off. The critical cut-off can be located between any two constraints. In each of the four rankings in (16) there are five positions where the critical cut-off can occur. These four rankings therefore represent a total of twenty rankings. Similarly, the ranking in (15b) also represents twenty possible rankings. This gives a total of forty rankings that are possible between the four constraints and the critical cut-off. These forty rankings represent all and only the possible ways in which [t, d]-deletion can interact with the following phonological context. I will not go through all forty possible rankings here. A discussion of all the possible rankings can be found in the Appendix at the end of this chapter. What I will do here is: (i) formulate the conditions that must be met for variation to be observed at all, (ii) list all the possible deletion patterns predicted under the analysis developed here (patterns that would result from at least one of these rankings), and (iii) mention some of the most important deletion patterns that are predicted to be impossible (that cannot result from any of these rankings).

1.2.3.1 Conditions for variation

In order to see what the general conditions are that must be met for variation to be observed, consider the pre-consonantal context as an example. There are two relevant candidates, namely the retention candidate [t, d] and the deletion candidate ∅. The retention candidate [t, d] violates only *Ct#C, and the deletion candidate ∅ violates only MAX. Variation between these candidates is only possible when neither of them is disfavored by a constraint ranked higher than the cut-off. This implies that variation will only be observed if the ranking ||Cut-off o {MAX, *Ct#C}|| is observed. The tableaux in (17) show the six different ways in which these two constraints and the cut-off can be ranked. Note that variation is only predicted when both constraints rank below the cut-off.

In (17a) and (17b) neither of the candidates violates a constraint higher than the cut-off. As a result both candidates will be accessed as outputs in this context. The ranking between MAX and the markedness constraint determines which of the candidates will be the more frequent variant. In (17c) and (17d) one candidate is disfavored by a constraint higher than the cut-off while the other is not. Since there is a candidate that is not disfavored by a constraint higher than the cut-off, no candidates that are disfavored by such a constraint will be accessed as possible output in these grammars. In (17e) and (17f) both candidates are disfavored by a constraint higher than the cut-off. In such a situation the language user has no choice but to select a candidate that is disfavored by a constraint ranked above the cut-off. However, only the single best candidate is selected when this happens. In order for variation to be observed in some context, it is necessary for both MAX and the markedness constraint that applies in that context to be ranked lower than the cut-off. In (18) this requirement is stated in general terms.

(17) a. **Variation 1**

/Ct#C/		MAX	*Ct#C
1	t		*
2	∅	*	

Output of EVAL

t *Ct# C

∅ MAX

Cut-off

b. **Variation 2**

/Ct#C/		*Ct#C	MAX
2	t	*	
1	∅		*

Output of EVAL

∅ MAX

t *Ct#C

Cut-off

c. **No variation 1**

/Ct#C/		MAX	*Ct#C
1	t		*
2	∅	*!	

Output of EVAL

t *Ct#C

Cut-off

∅ MAX

d. **No variation 2**

/Ct#C/		*Ct#C	MAX
2	t	*!	
1	∅		*

Output of EVAL

∅ MAX

Cut-off

t *Ct#C

e. **No variation 3**

/Ct#C/		MAX	*Ct#C
1	t		*
2	∅	*!	

Output of EVAL

Cut-off

t *Ct#C

∅ MAX

f. **No variation 4**

/Ct#C/		*Ct#C	MAX
2	t	*!	
1	∅		*

Output of EVAL

Cut-off

∅ MAX

t *Ct#C

(18) **Necessary conditions for variation in each context**

a. General condition: ||Cut-off o MAX||

b. Context specific conditions:

(i) Pre-Consonantal: ||Cut-off o *Ct#C||

(ii) Pre-Vocalic: ||Cut-off o *Ct#V||

(iii) Pre-Pausal: ||Cut-off o *Ct##||

In addition to the requirements in (18) there is, of course, also one universally fixed ranking that needs to be taken into consideration – the constraint against [t, d] in pre-consonantal context universally outranks the constraints against [t, d] in the other two contexts. The way in which the conditions in (18) interact with this universal ranking determines the different possible deletion patterns. These patterns are discussed in more detail in the next section.

1.2.3.2 Possible deletion patterns

In (19) below I list all the deletion patterns that are predicted as possible by the analysis developed above. I do not motivate here that these are all and only the possible patterns. Nor do I give the rankings that are necessary for each of the patterns. This discussion can be found in the Appendix at the end of this chapter. In table (19) "D" stands for "categorical deletion", "R" for "categorical retention", "D > R" for "more deletion than retention" and "R > D" for "more retention than deletion". In order to make identifying patterns easier, I also shade all cells in which more deletion than retention is observed – i.e. both "D" and "D > R" cells.

(19) **Possible deletion patterns**

a. **No-variation**

Pre-Consonantal	Pre-Vocalic	Pre-Pausal
D	D	D
D	D	R
D	R	D
D	R	R
R	R	R

b. **Variation in all contexts**

Pre-Consonantal	Pre-Vocalic	Pre-Pausal
D > R	D > R	D > R
D > R	D > R	R > D
D > R	R > D	D > R
D > R	R > D	R > D
R > D	R > D	R > D

c. **Variation only in some contexts**

Pre-Consonantal	Pre-Vocalic	Pre-Pausal
D	D	D > R
D	D > R	D
D	D > R	D > R
D	D	R > D
D	D > R	R > D
D	R > D	D
D	R > D	D > R
D	R > D	R > D

Of the predicted patterns in (19a), none are actually observed. However, all of these seem to be very reasonable and likely. The last pattern, with categorical retention in all three contexts, would be a very conservative dialect. This is at least the normative

dialect of standard American English, and this pattern might even be observed in very careful hyper-articulated speech.

The first pattern, with categorical deletion everywhere, is at the other endpoint of the spectrum. Of course, if the [t, d] is always deleted then re-lexicalization will occur – i.e. the next generation of learners will acquire underlying forms without the final [t, d]'s.[11] A dialect like Jamaican English, with high deletion rates in all three contexts (see (1) above), might be en route to this point. In fact, Patrick (1991) argues that re-lexicalization might indeed already have occurred for many words in this dialect.

Several of the patterns in (19b) are actually attested. For the deletion patterns of the different dialects, see (1) and (2) above. An important characteristic that all of the predicted patterns have in common is that pre-consonantal context shows at least as much deletion as the other two contexts – that is, there is no pattern where pre-consonantal context prefers retention and one of the other contexts prefers deletion. Under the licensing by cue analysis this is to be expected.

According to Guy (1980:27) the English spoken by white Americans in both New York and Philadelphia has categorical deletion in pre-consonantal context and variable deletion in the other contexts. Unfortunately Guy reports only the VARBUL factor weights for these dialects. Pre-consonantal context has a factor weight of 1.0 which translates into categorical deletion. The deletion rates for the other two contexts cannot be determined from their factor weights. We therefore know that these two dialects fall into one of the patterns in (19c), but we do not know in which specific one.

---

[11] This is true only for monomorphemes. Since the past tense of verbs that end on vowels will still be marked by a /-d/ suffix, the past tense of all verbs is still very likely to contain this /-d/-suffix.

1.2.3.3 Impossible deletion patterns

Perhaps more instructive are the deletion patterns that are predicted as impossible because they can never result from any of the possible rankings. If any of these patterns are actually encountered, it would count as strong evidence against the analysis developed above. Of the patterns that are predicted as impossible, none are actually attested to my knowledge.

One group of deletion patterns that are predicted as impossible, are patterns where pre-vocalic or pre-pausal contexts show more deletion than pre-consonantal context. The non-existence of these patterns is reasonable and expected. Pre-consonantal context is the weakest sponsor for [t, d]. We would therefore not expect to see more [t, d]'s retained in this context than in the other contexts that are more robust sponsors for [t, d]. A sample of these patterns is listed in (20).

(20) **Impossible patterns: More retention in pre-consonantal context than in the other contexts**

Pre-Consonantal	Pre-Vocalic	Pre-Pausal
R	D	D
R	D	R
R	R > D	R
R	R	R > D
R > D	R > D	D
R > D	D	R > D
R > D	D > R	R
R > D	R	D > R
R > D	D	D
D > R	R > D	D
D > R	D	R > D
D > R	D	D

The impossibility of these patterns follows from the universally fixed ranking between the markedness constraints in (3b): ||*Ct#C o {*Ct#V, *Ct##}||. Because of this ranking retention of [t, d] in pre-consonantal context will always be more marked than retention in either pre-vocalic or pre-pausal context. Since deletion is motivated by the drive to become less marked, the drive to delete will always be strongest in pre-consonantal context, and this context will therefore always show at least as much deletion as the other two contexts.

The exclusion of these patterns is a very desirable prediction of the analysis developed here, and can serve as a strong argument in favor of this analysis. However, there is also one group of patterns that are predicted as impossible even though they seem quite reasonable. I will discuss these patterns next.

Deletion patterns with variation in pre-consonantal context and categorical retention in pre-vocalic and/or pre-pausal position are predicted as impossible under the analysis developed above. I list a few examples of these patterns in (21).

Why are these patterns impossible in the account developed here? In order for variation to be observed in pre-consonantal position, the ranking ||Cut-off o {MAX, *Ct#C}|| is required (see the variation conditions in (18)). However, because the markedness constraint against [t, d] in pre-consonantal context universally outranks the constraints against [t, d] in pre-vocalic and pre-pausal context (see (3b)), we have by transitivity of constraint ranking also the ranking ||Cut-off o {MAX, *Ct#V, *Ct##}||. The variation conditions in (18) are therefore also met for pre-pausal and pre-vocalic context. The implication is that variation in pre-consonantal context is always accompanied by variation in pre-pausal and pre-vocalic context.

(21) **Impossible patterns: Variation in pre-consonantal context, and categorical retention in pre-pausal and/or pre-vocalic position**[12]

Pre-Consonantal	Pre-Vocalic	Pre-Pausal
D > R	R	R
D > R	R	R > D
D > R	R	D > R
D > R	R	R
D > R	R > D	R
D > R	D > R	R
R > D	R	R
R > D	R	R > D
R > D	R > D	R

These patterns all seem reasonable. In all of these patterns the less robust pre-consonantal sponsoring context of [t, d] shows at least as much deletion as the more robust pre-pausal and pre-vocalic sponsoring contexts. Although this prediction of the analysis seems potentially problematic, to the best of my knowledge no dialect has been reported to show any of these patterns. At the present time, I will therefore only acknowledge this as a falsifiable prediction of the analysis.

1.2.4 Two outstanding questions

In this section I will briefly discuss two outstanding questions. The analysis that I have developed above assumes that only [t, d] deletes. What about other consonants that

[12] There are more patterns that would fit this general description, for instance:

Pre-Consonantal	Pre-Vocalic	Pre-Pausal
D > R	R	D
R > D	R	D > R

These patterns all show less deletion in pre-consonantal than pre-vocalic or pre-pausal context. They are therefore excluded under (20) and are not problematic at all.

appear word-finally in consonant clusters? I address this question in §1.2.4.1. The analysis also assumes that word-final [t, d] only deletes if it is part of a consonant cluster. What about word-final [t, d] that follows directly after a vowel? This question is discussed in §1.2.4.2.

1.2.4.1 Only [t, d]?

The markedness constraints that drive the deletion have been formulated above to refer specifically to [t, d]. Consequently, consonants other than [t, d] that occur as final members in word-final clusters will not violate these constraints. The prediction is therefore that only [t, d] will delete. This seems unlikely. If [t, d] deletes from words like *mist*, then it seems reasonable to expect that [k] might delete from a word like *whisk*, that [p, b] might delete from words like *wasp* or *bulb*, that [θ] might delete from words like *filth*, etc. In the rest of this section I will limit discussion to the other stop consonants, i.e. [k, g, p, b]. The reason for this is that the markedness constraints were formulated in terms of the cues necessary to identify [t, d]. Since at least some of the cues for the identification of fricatives and sonorants differ from those used in the identification of stops, it seems reasonable that these constraints will not apply to non-stop consonants.

With regard to the non-coronal stop consonants, there is in fact acknowledgement in the literature that they do delete in the same contexts in which [t, d] deletes (Guy, 1980:1, Labov *et al*., 1968:131-133, Wolfram, 1969:50). We therefore have to consider the implication of this fact for the analysis developed for [t, d]-deletion. Even though it is acknowledged that stops other than [t, d] also delete form word-final clusters, no data

exist on the patterns associated with deletion of these other consonants.[13] We can therefore only speculate on how the deletion of these consonants will pattern. In the rest of this section I will consider two possibilities about how deletion of non-coronals might pattern, and show how we could account for these patterns in the analysis developed above for [t, d]-deletion.

The most straightforward possibility is that the deletion process does not distinguish between the stops in terms of place of articulation (i.e. [k, g, p, b][14] are subject to the same rates of deletion as [t, d] in the different contexts). It is very unlikely that this will be the case. However, if it were to be the case, it would be very easy to adapt the analysis developed above to accommodate the non-coronals. We can simply redefine the markedness constrains so that they are constraints on where stop consonants can occur rather than constraints on where coronal stops can occur. Everything else can be left as is.

A more likely scenario is that the non-coronals are subject to lower deletion rates than the coronals. Ohala (1990) conducted perception experiments in which he presented listeners with [VC_1-C_2V]-sequences. The tokens in the experiments differed in the length of the silence gap between C_1 and C_2. When this silence gap dropped below 100 ms in duration, listeners tended to perceive only one consonant rather than two. The single consonant percept was usually identical to C_2 rather than to C_1. This perceptual

---

[13] The reason for this is undoubtedly that there is only a very small number of word-final consonant clusters that end in non-coronals (Fudge, 1969), and consequently there are very few words that end in these clusters. In order to identify with confidence patterns that might arise in the deletion of non-coronal stops we need a large number of tokens in which these consonants occur in the relevant context. However, since these kinds of tokens are so scarce, it is practically impossible to collect enough data on the deletion of these consonants.

[14] Actually, [g] never appears as final member of a word-final consonant cluster – see Fudge (1969). [g] is therefore included in the discussion simply for the sake of completeness.

phenomenon is expected under the perceptual licensing approach to neutralization assumed above. Many of the cues for the perception of a consonant are carried in the consonantal release. As the duration between C_1 and C_2 decreases, there is less and less time for the release cues of C_1 to be realized. At a certain point these cues are so weakly realized that C_1 is not perceived at all.[15]

Kingston and Shinya (2003, see also Kingston, to appear) replicated these findings with stop consonants. However, they also extended the results in an interesting manner. They distinguished between the places of articulation, and found a difference between coronal and non-coronal stops. They calculated the likelihood that a coronal stop followed by a non-coronal will be identified as identical to the non-coronal.[16] They also calculated similar statistics for the non-coronals – i.e. the likelihood that a labial followed by a non-labial will be identified as the non-labial, and the likelihood that a velar followed by a non-velar will be identified as the non-velar. They found that the likelihood of identifying a coronal incorrectly is larger than the likelihood of identifying a labial or a velar incorrectly. Based on this result, Kingston and Shinya conclude that, in [VC_1-C_2V]-sequences, C_1 is more likely to delete if it is a coronal than if it is a non-coronal.

In all of these studies, the coronals did not appear as final member of a consonant cluster – they were directly preceded by a vowel. However, it is reasonable to assume that the same pattern would emerge even if C_1 in the sequence were preceded not by a

---

[15] See also Repp (1978, 1982, 1983). He conducted experiments similar to those of Ohala and he found very similar results. However, unlike Ohala who used naturally produced stimuli that therefore could contain releases, Repp used synthesized stimuli. For the most part his stimuli did not include releases – i.e. the only cues to place of articulation were in the formant transitions.

[16] If C_1 in a [VC_1-C_2V]-sequence is identified as identical to C_2 it can be interpreted as deletion of C_1, especially in light of the fact that English does not tolerate geminates.

vowel but by another consonant. Based on these results we can therefore expect that non-coronals will be subject to lower deletion rates than coronals.

If this turns out to be true, there are several ways in which the analysis developed for [t, d]-deletion above can be extended to account for the difference between coronals and non-coronals. I mention only the one of the more feasible options here. It is possible to reformulate the markedness constraints used above so that they do not refer to [t, d] specifically, but rather to all stop consonants. We can then define MAX constraints indexed to the different places of articulation – i.e. $\text{MAX}_{[\text{Cor}]}$ (violated by deletion of a coronal) and $\text{MAX}_{[\text{Non-Cor}]}$ (violated by deletion of a non-coronal). By ranking $\text{MAX}_{[\text{Non-Cor}]}$ higher than $\text{MAX}_{[\text{Cor}]}$ we can account for the lower deletion rates of non-coronals. Deletion of a non-coronal will violate a higher ranked faithfulness constraint and will therefore be more costly.

It does not seem unreasonable to define MAX constraints that refer to place of articulation. De Lacy (2002) argues that there is a universal tendency to be more faithful to the more marked elements on a markedness scale. It is generally accepted that coronals are less marked than non-coronals – see the place markedness hierarchies in *inter alia* de Lacy (2002), Gnanadesikan (1996), Jakobson (1968), Lombardi (2001), Prince (1998), and also the discussion on [sCvC]-forms in Chapter 6 §3.1.2.1. If we do have MAX constraints indexed to place of articulation, we would therefore expect the constraints referring to non-coronals to rank higher than the constraint referring to coronals.

1.2.4.2 What about [t, d] preceded by a vowel?

The definition of the markedness constraints that drive the deletion of [t, d] includes reference to a preceding consonant (see (3a)). As a consequence these markedness

constraints will not be violated by a post-vocalic word-final [t, d]. From this follows that [t, d] will only delete from a consonant cluster. However, we know that this is not true. Even [t, d] that occurs in simplex codas does delete. Fasold (1972:41) reports that the final [d] in a word like *applied* is often not pronounced in AAE.[17] Mees reports that post-vocalic [t] in Cardiff English is subject to a weakening process. It most often glottalizes, but it is also deleted rather frequently (Mees, 1987). This is probably true of all glottalizing dialects of English. The studies by Repp (1978, 1982, 1983), Ohala (1990) and Kingston and Shinya (2003) discussed in the previous section also show that consonants can delete from simplex codas – all of these studies showed that C_1 can delete from [VC_1-C_2V]-sequences.

Since we know that post-vocalic word-final [t, d] is also subject to deletion, a more complete analysis of this phenomenon will take this fact into account. Unfortunately, other than acknowledgement of the fact that deletion also applies in this context, very little information is available on the actual deletion patterns associated with post-vocalic [t, d]. The literature on [t, d]-deletion reports nearly exclusively on deletion of [t, d] in post-consonantal context. We are therefore again forced to speculate about how [t, d]-deletion will pattern in post-vocalic context.

It is most likely the case that [t, d] will delete much less frequently in post-vocalic than post-consonantal position. There are at least two reasons for this. The literature on [t, d]-deletion devote a lot of attention to the influence of the preceding context on the rate of [t, d]-deletion. (Section §2 of this chapter is also devoted to that.) The general finding

---

[17] Schouten also reports that final [t, d] in Dutch deletes even when it is preceded by a vowel (Schouten, 1982:285).

is that [t, d] deletes more the more similar the preceding segment is to [t, d] – i.e. it deletes more if the preceding segment is a stop (*pact*) than if it is a fricative (*drift*), it deletes more if the preceding segment is a coronal (*banned*) than if it is a non-coronal (*crammed*), etc. Vowels are undoubtedly more different from [t, d] than any consonant. Based on this we would expect lower deletion rates after a vowel than after any consonant.

Also under the licensing by cue approach we would expect less deletion in post-vocalic than in post-consonantal position. In the discussion of the markedness constraints in §1.2.1 above, I have focused only on the perceptual cues that are present in the context following a consonant. However, there are also cues available in the context preceding the consonant. In the same manner that consonants are cued by formant transitions into following vowels, they are also cued by formant transitions from preceding vowels. There are therefore more cues available to the identity of a post-vocalic than a post-consonantal consonant.

Assuming that this speculation is correct – i.e. that (i) post-vocalic [t, d] also deletes, but (ii) at a lower rate than post-consonantal [t, d] – how can the analysis developed above be changed to incorporate this? The most obvious manner is to formulate three more markedness constraints. For each of the constraints in (3a) above, there will then be a counterpart that applies in post-vocalic position – i.e. a constraint against [t, d] in the context V __ # C, in the context V __ # V, and in the context V __ ##. Since these contexts each contain an extra cue to the identity of [t, d], they sponsor [t, d] more robustly. They will therefore rank lower than the constraints that apply in post-consonantal position. Since the post-vocalic constraints rank lower, candidates that

violate them will be less marked than candidates that violate the post-consonantal constraints. The prediction would then be less deletion in post-vocalic than in post-consonantal context.

1.3 An alternative explanation

The analysis that I have developed above depends on the assumption that [t, d]-deletion is driven by licensing constraints rather than ordinary syllabic well-formedness constraints. [t, d] deletes when the cues for its perception are not robust enough. An alternative account is one that depends on syllabic well-formedness. [t, d] deletes from word-final clusters because a complex coda is more marked than a simplex coda – deletion of [t, d] is then driven by a constraint like *COMPLEX. Kiparsky (1993) and Reynolds (1994:119-137) (following Kiparsky) take this approach. In this section I will briefly review evidence against this *COMPLEX-approach.

In all the dialects of English that has been studied, pre-consonantal context shows more deletion than pre-vocalic context. This seems to be a strong generalization that should be formally captured in our analysis – i.e. the analysis should exclude the existence of a dialect with equal variable deletion rates in these two contexts. In the analysis that I developed in §1.2 this is indeed predicted to be the case. In this analysis there are different markedness constraints that drive the deletion of [t, d] in pre-vocalic and pre-consonantal context, namely *Ct#V and *Ct#C. The difference in deletion rates between the contexts then follows from the fact that non-deletion violates different markedness constraints in the different contexts.

However, the *COMPLEX-approach is different. Under this approach a grammar with variable deletion that does not distinguish between pre-vocalic and pre-consonantal

context is possible (i.e. a dialect with equal variable deletion in pre-vocalic and pre-consonantal context). In this approach there is only one constraint that drives deletion in both pre-consonantal and pre-vocalic context. The difference in deletion rates between these contexts therefore has to be explained in a different manner. In this approach the difference is attributed to re-syllabification. The assumption is that a *COMPLEX-violation in pre-vocalic context can be avoided in two ways – either by deletion or by re-syllabifying the [t, d] across the word-boundary into the onset of the following syllable. A form such as *west end* can then be pronounced in three ways: (i) with a final consonant cluster [wɛst.ɛnd], (ii) with re-syllabification of the final [t] form *west* into the onset of the following syllable [wɛs.tɛnd], or (iii) with deletion of the final [t] from *west* [wɛs.ɛnd]. In pre-consonantal context, however, re-syllabification is not available as a way in which to avoid a *COMPLEX-violation. A form such as *west bank* can be pronounced in only two ways: (i) with a final consonant cluster in *west* [wɛst.bæŋk], or (ii) with deletion of the final [t] from *west* [wɛs.bæŋk]. Re-syllabification [wɛs.tbæŋk] is not available as an option.[18] Pre-vocalic context then shows higher retention rates since it can avoid violation of *COMPLEX without having to delete the final [t, d].

[18] In some pre-consonantal contexts re-syllabification is available. For instance, in an utterance such as *west rock* it is possible to re-syllabify the final [t] of *west* into the following syllable [wɛs.trɔk]. However, this is only possible for a small subset of consonants that could start the following word. [t, d] can only be followed by a small set of consonants in the onset of a syllable. Since final [t, d] can always re-syllabify in pre-vocalic position and can only re-syllabify in front of a small number of consonants, the assumption is that the small number of pre-consonantal re-syllabifications will not have an appreciable effect on the overall deletion pattern.

See also Labov (1997) who studied the speech of two English speakers in detail. He reports that there is very little evidence that these speakers re-syllabify across a word-boundary when the following word starts in a consonant.

The lower deletion rate in pre-vocalic position therefore crucially depends on the assumption that re-syllabification is allowed. If re-syllabification were not allowed, then deletion would be the only way to avoid a *COMPLEX-violation both in pre-vocalic and pre-consonantal contexts. Re-syllabification does not come for free. A form that re-syllabifies final [t, d] across a word boundary violates a constraint from the ALIGN family (McCarthy and Prince, 1993b). In a form such as [wɛs.tɛnd] the right edge of the morphological word is not aligned with the right edge of any prosodic category.[19] In order for a re-syllabification candidate like [wɛs.tɛnd] to be available as a variant pronunciation of *west end* it is therefore necessary that the form [wɛs.tɛnd] is not disfavored by any constraint ranked higher than the critical cut-off. If ALIGN ranks above the critical cut-off, then [wɛs.tɛnd] would not be allowed as a pronunciation, and there would be no difference between pre-vocalic and pre-consonantal context. This is illustrated in the tableau in (22). This tableau represents a dialect with variable deletion in both pre-vocalic and pre-consonantal context. This requires the ranking ||Cut-off ο {MAX, *COMPLEX}|| so that neither deletion nor retention is disfavored by a constraint ranked higher than the cut-off.

In pre-consonantal context there are two candidates that are not disfavored by any constraints ranked higher than the cut-off, namely the deletion candidate and the retention candidate. Of these, the retention candidate is rated better by EVAL and is therefore expected to be the more frequently observed variant. The same is true in pre-vocalic context. But now there is no difference between these two contexts – since the re-

[19] The exact formulation of this ALIGN constraint is not relevant here. It can be a constraint that requires the morphological word to be right-aligned with a syllable, a prosodic word or a foot. All that is relevant here is that the re-syllabification candidate violates some ALIGN constraint.

syllabification candidate is not available as a variant in pre-vocalic context. The retention candidates from both contexts are also both equally marked (since both violate only *COMPLEX), so that the drive to delete is equally strong in both contexts. Consequently, it is predicted that both contexts should show the same deletion rate.

(22) **No difference between pre-consonantal and pre-vocalic context**

			DEP	ALIGN	MAX	*COMPLEX
/Ct#C/	1	Ct.C				*
	2	C∅.C			*	
	3	CVt.C	*!			
/Ct#V/	1	Ct.V				*
	3	C.tV		*!		
	2	C∅.V			*	
	4	CVt.V	*!			

Output of EVAL

Pre-Consonantal	**Pre-Vocalic**
Ct.C *COMPLEX	Ct.V *COMPLEX
C∅.C MAX	C∅.V MAX
Cut-off	
CVt.C DEP	C.tV ALIGN
…	…

Under the *COMPLEX-approach it is therefore predicted that dialects with equal variable deletion in pre-consonantal and pre-vocalic contexts are possible. This is a problematic prediction. No such dialects have been reported in the literature, and the standard assumption is that pre-consonantal context will always show higher deletion

rates than pre-vocalic context. There is a way in which to prevent the *COMPLEX-analysis form allowing dialects like those in (22). What we want to exclude are dialects with variable deletion in pre-consonantal and pre-vocalic contexts, but that do not allow re-syllabification. Variable deletion implies that both MAX and *COMPLEX rank below the cut-off, and the non-availability of re-syllabification implies that ALIGN ranks above the cut-off. What we need is to stipulate that the ranking ||Cut-off o {MAX, *COMPLEX}|| implies the ranking ||Cut-off o ALIGN||. This stipulation would successfully exclude a grammar like that in (22). However, this is a very arbitrary stipulation without any conceivable substantive motivation.

The cue-based analysis that I developed in §1.2 above is fundamentally different. Under this analysis any dialect with variable deletion in both pre-consonantal and pre-vocalic contexts will necessarily show more deletion in pre-consonantal context. The reason for this is that non-deletion in pre-consonantal and pre-vocalic contexts violates different constraints, and that there is a universal ranking between these two constraints such that non-deletion in pre-consonantal context will always be more marked than non-deletion in pre-vocalic context, i.e. ||*Ct#C o *Ct#V||. The ranking ||*Ct#C o *Ct#V|| precludes the existence of a dialect with equal variable deletion in pre-consonantal and pre-vocalic context. Non-deletion in a pre-consonantal context violates a higher ranking markedness constraint, so that the drive to delete is stronger in this context than in pre-vocalic context. If both pre-consonantal and pre-vocalic contexts show variable deletion, pre-consonantal context will always have a higher deletion rate.

2. The preceding phonological context

The nature of consonant preceding the [t, d] also influences the rate of [t, d]-deletion. The general finding is that certain preceding consonants (such as the sibilants) are associated with higher [t, d]-deletion rates than other segments (such as the liquids). In this section I will analyze the effect of the preceding segment on [t, d]-deletion within the rank-ordering model of EVAL. The analysis that I develop builds on the claims by Guy (1994) and Guy and Boberg (1997) that we are dealing with an Obligatory Contour Principle (OCP) effect. The idea is that [t, d] deletes in order to avoid two contiguous consonants that are too similar. The more similar the preceding consonant is to [t, d], the more likely [t, d] is to delete.

The rest of this section is structured as follows: In §2.1 I present the data on the influence of the preceding consonant on the rate of [t, d]-deletion. Section §2.2 then contains the actual analysis of these data within the rank-ordering model of EVAL.

2.1 The data

In the literature on [t, d]-deletion the consonants that precede the [t, d] are usually classified in terms of gross manner features – i.e. we find different combinations of sound classes such as sibilants, non-sibilant fricatives, stops, nasals, and liquids. There is some variation between dialects, but the general finding is that these consonant classes can be ordered as follows in terms of declining [t, d]-deletion rates: sibilants > stops > nasals > non-sibilant fricatives > liquids (Labov, 1989:90). No completely satisfactory explanation could ever be given for this specific pattern. Most often it was interpreted as a sonority effect – the less sonorous the segment preceding the [t, d], the more likely [t, d] is to delete. Although this is generally true, there are several exceptions to this generalization.

The sibilants (as fricatives) are certainly more sonorous than the stops – why then are they associated with higher deletion rates than the stops? The nasals are also more sonorous than the fricatives – why do nasals then show higher deletion rates than the non-sibilant fricatives?

Labov classifies the preceding context as "a relatively weak constraint" (Labov, 1989:90). This classification is intended to show that the preceding context is a less accurate predictor of the likelihood of [t, d]-deletion than the following context. If we know that the mean [t, d]-deletion rate before all vowels is *x* percent, then we can expect that the mean [t, d]-deletion rate before some specific vowel will also be roughly *x* percent. However, this is not true with regard to the preceding context. If we knew that the mean [t, d]-deletion rate after all nasals is *y* percent, then we cannot necessarily assume that the mean [t, d]-deletion rate after any specific nasal will be *y* percent. There is more scatter in the data on the preceding context than in the data on the following context.

Rather than interpreting this as a consequence of the fact that the preceding context has a weaker influence on the likelihood of [t, d]-deletion relative to the following context, Guy (1994) and Guy and Boberg (1997) argue that the greater degree of scatter in the data on the preceding context should be taken as an indication of the fact that the preceding context has not been partitioned correctly. We should define different groupings of consonants than those traditionally used in the literature. They argue that the relevant classes should be consonants that share the same features with [t, d]. [t, d] is uniquely identified by the three features [+coronal, -continuant, -sonorant]. The consonants that precede [t, d] should therefore be classified according to which of these

features they share with [t, d]. Guy and Boberg analyze a corpus based on the speech of three Philadelphians according to the classes defined by this metric.[20] The results of this analysis are given in (23) (Guy, 1994:145, Guy and Boberg, 1997:155).

(23) **The influence of the preceding context on [t, d]-deletion rates in the Guy and Boberg (1997) corpus**

Shared features	Consonants	*n*	% deleted
?	[r]	86	7
[-cont]	[m, ŋ]	9	11
[-son]	[f, v]	45	29
[+cor]	[l]	182	32
[-cont, -son]	[k, g, p, b]	136	37
[+cor, -cont]	[n]	337	46
[+cor, -son]	[s, š, z, ž]	276	49
[+cor, -cont, -son]	[t, d]	–	–

In the table in (23) the contexts are ordered such that contexts with higher deletion rates occur lower in the table. The deletion rates between some contexts are clearly different – for instance, after [s, š, z, ž] 49% of [t, d] deletes but after [f, v] only 29% of [t, d] deletes. However, for some contexts it is not so clear whether their deletion rates are really different – for instance, is the 49% deletion rate after [s, š, z, ž] really different from the 46% deletion rate after [n]? In order to answer these questions, I calculated a χ^2-statistic for every pair of contexts. The results of this calculation are given in (24). The

[20] Unfortunately, Guy and Boberg report no further details about their corpus. We therefore do not know anything more about the three individuals on whom this corpus is based.

top number for every context is the $\chi^2(1)$ statistic for that pair of contexts, and the bottom number is the *p*-value. Pairs with *p*-values smaller than 0.05 are italicized, and pairs with larger *p*-values are underscored.

(24) **$\chi^2(1)$ and *p*-values for context pairs in the Guy and Boberg corpus**[21]

	[+cor, -son] **[s, š, z, ž]**	**[+cor, -cont]** **[n]**	**[-cont, -son]** **[k, g, p, b]**	**[+cor]** **[l]**	**[-son]** **[f, v]**	**[-cont]** **[m, ŋ]**	**?** **[r]**
[+cor, -son] **[s, š, z, ž]**							
[+cor, -cont] **[n]**	0.94 0.33						
[-cont, -son] **[k, g, p, b]**	*16.8* *< 0.000*	*11.69* *< 0.000*					
[+cor] **[l]**	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	1.42 0.23				
[-son] **[f, v]**	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	*3.98* *0.046*	0.73 0.394			
[-cont] **[m, ŋ]**	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	*14.71* *< 0.000*		
? **[r]**	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	*> 10^{307}* *< 0.000*	0.23 0.629	

[21] The $\chi^2(1)$ statistic for a pair of contexts was calculated as follows: Let the two contexts be $Cont_1$ and $Cont_2$, such that $Cont_1$ has a higher deletion rate than $Cont_2$. The observed values were taken to be the actual number of deletions and preservations of $Cont_1$. The expected values were the deletion and preservation rates that would be observed in $Cont_1$ had $Cont_1$ shown the same deletion rate as $Cont_2$. Consider the [s, š, z, ž] and [n] contexts as an example. [s, š, z, ž] has a deletion rate of 49%, and [n] has a deletion rate of 46%. The observed values were therefore taken to be the actual deletion and retention values of [s, š, z, ž], while the expected values were calculated assuming that [s, š, z, ž] had also shown a 46% deletion rate.

	Observed (49% deletion)	Expected (46% deletion)
Deletion	135	126.96
Retention	141	149.04

What can we conclude based on these comparisons? First consider the contexts that share only one feature with [t, d]. [t, d] deletes more after both [+coronal] and [-sonorant] than after [-continuant] – i.e. more deletion after [l, f, v] than after [m, ŋ]. Co-occurrence of two sounds that agree in place or sonorancy is therefore avoided more strongly than co-occurrence of two sounds that agree in continuancy.[22] Based simply on comparison of [+coronal] and [-sonorant] contexts, it is not possible to order these two contexts in terms of deletion rates – i.e. it is not possible to determine whether [t, d] deletes more following [l] or following [f, v]. Although [+coronal] context has a numerically higher deletion rate, the difference in deletion rate between these two contexts is not statistically significant. However, based on universal tendencies we would expect that co-occurrence of two homorganic consonants will be avoided more than co-occurrence of two consonants that agree in sonorancy. Consonantal co-occurrence patterns in the Semitic languages, for instance, are defined in terms of place of articulation – there are restrictions on the co-occurrence of homorganic consonants but heterorganic consonants co-occur freely (Frisch *et al*., 2004, Greenberg, 1950, McCarthy, 1994, Pierrehumbert, 1993). There are also two bits of indirect evidence that [+coronal] does contribute more towards [t, d]-deletion than [-sonorant]. First, [t, d] deletes more frequently after [n] than after [k, g, p, b]. These two contexts share the [-continuant] feature, and differ only in that [n] is [+coronal] while [k, g, p, b] are [-sonorant].

---

[22] This is not unexpected. We know from consonantal co-occurrence patterns in other languages that co-occurrence of segments that agree in place is avoided more than co-occurrence of segments that do not agree in place of articulation. This is true at least for Arabic (Frisch *et al*., 2004), English, French and Latin (Berkley, 1994a, 1994b, 2000). The fact that the [+coronal] context shows more deletion than [-continuant] context is therefore expected. Similarly, it is not unexpected that co-occurrence of two consonants that agree in sonorancy is avoided more than co-occurrence of two consonants that agree in continuancy. At least in Arabic, co-occurrence of coronals that agree in sonorancy is avoided more than co-occurrence of coronals that agree in continuancy.

Secondly, [k, g, p, b] have a significantly higher deletion rate than [f, v], but not than [l]. These two facts taken together with the fact that [l] has a numerically higher deletion rate than [f, v] suggests that [+coronal] does contribute more toward [t, d]-deletion than [-sonorant].

Now we can consider the contexts that share two features with [t, d]. The first observation that can be made is that both [s, š, z, ž] and [n] show higher deletion rates than [p, b, k, g]. Co-occurrence of consonants that share [+coronal] and either [-sonorant] or [-continuant] is avoided more than co-occurrence of two consonants that share [-sonorant] and [-continuant]. However, although [s, š, z, ž] has a numerically higher deletion rate than [n], this difference is not statistically significant. The expectation would be that [s, š, z, ž] will have a higher deletion rate – since these segments share both [+coronal] and [-sonorant] with [t, d], and we have seen just above that these two features contribute more to deletion than [-continuant]. Although there is no direct evidence that [s, š, z, ž] have a higher deletion rate than [n], there is some indirect evidence that suggests this to be the case. If we calculate the probability that the deletion rate in each of these two contexts differ from chance (i.e. from 50%), there is a strong indication that [s, š, z, ž] is associated with a higher deletion rate than [n]. Under the binomial distribution, the likelihood that the deletion rate after [s, š, z, ž] is really 50% is 0.38. However, the likelihood that the deletion rate after [n] is really 50% is only 0.08.

Now we can connect the contexts that share two features with [t, d] and the contexts that share only one feature with [t, d]. The first thing that can be noted is that [t, d] deletes more after both [s, š, z, ž] and [n] than after any of the consonants that share only one feature with [t, d]. All that we have to determine is the relationship of [k, g, p, b]

to the segments that share only one feature with [t, d]. The [k, g, p, b] context has a higher numerical deletion rate than the [l] context. However, this difference is not significant. It is therefore not clear whether co-occurrence of two consonants that share both [-continuant] and [-sonorant] is avoided more or whether co-occurrence of two consonants that share only [+coronal] is avoided more. However, the [k, g, p, b] context has a higher deletion rate than both the [f, v] and [m, ŋ] contexts. This makes sense – sharing both [-continuant] and [-sonorant] is avoided more than sharing only one of these features.

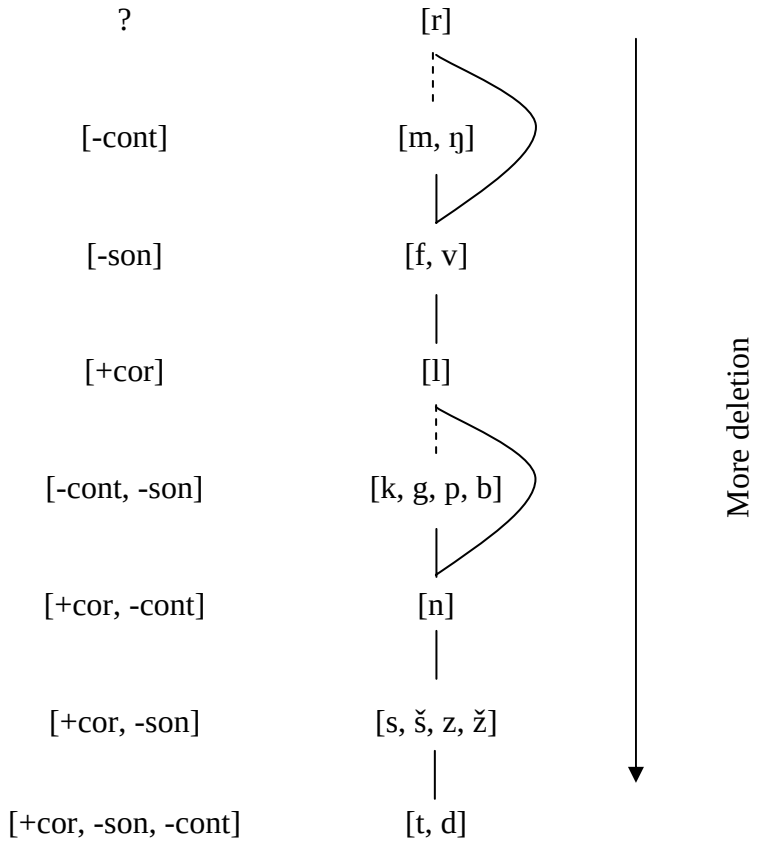
What remains now is to consider preceding [t, d] and preceding [r]. There are no monomorphemes in the English lexicon that end in one of the sequences [-tt], [-td], [-dd], or [-dt]. There are also no such bi-morphemic forms – the past tense suffix used with verbs that end in /t, d/ is [-əd]. It is therefore not possible to determine the [t, d]-deletion rate after [t, d]. However, the fact that such forms are completely absent from English suggests that English does not tolerate them at all. Had such forms existed, they would probably have been associated with the highest deletion rate, most likely a 100% deletion rate. This also makes sense in the general pattern of [t, d]-deletion observed in the table in (24). [t, d] deletes more after segments that share two features with [t, d] than after segments that share only one feature with [t, d]. We can therefore expect that [t, d] will delete even more after [t, d] with which it shares all three of the relevant features.

Lastly, consider the preceding [r] context. All contexts except for [m, ŋ] have higher deletion rates than [r]. If we take [r] to be a coronal liquid like [l], then [r] should show the same high deletion rate of [l]. Guy and Boberg take the low deletion rate after [r] as an indication that post-vocalic [r] is usually not pronounced at all or is realized only as

[r]-coloring on the preceding vowel. When this happens the [t, d] is preceded by a vowel and we would expect a low [t, d]-deletion rate (see §1.2.4.1 and §2.2.4.1 for a discussion of [t, d]-deletion in a post-vocalic context). According to Guy and Boberg, the post-vocalic [r] in their data were at least sometimes pronounced as a consonant. In these contexts we would then expect higher deletion rates, similar to those in the [l]-context.

In (25) I give a graphic representation of the order between the preceding contexts. A context that appears higher is associated with a lower [t, d]-deletion rate. A solid line between two contexts means that there is either direct or indirect evidence that the deletion rates between the two contexts are different. A broken line means that the deletion rates differ, but that strong evidence for the difference is lacking.

(25) **Order between different preceding contexts in the Guy and Boberg corpus in terms of [t, d] deletion rates**



Overall we have pretty strong evidence that the contexts are actually ordered as shown in (25). The two orderings that are less sure, are those between [k, g, p, b] and [l], and that between [m, ŋ] and [r]. About the latter of these two, we cannot conclude much. Guy and Boberg do not distinguish between [r] that is pronounced consonantally and [r] that is either not pronounced or that is pronounced as [r]-coloring on the preceding vowel. It is therefore really only about the order between [k, g, p, b] and [l] that the available data are not conclusive. Abstracting away from the uncertainties about these two orderings, I will assume that (25) represents the actual order between preceding contexts in terms of [t, d]-deletion rates, and the analysis that I will develop in §2.2 below will account for this order.

As mentioned at the beginning of this section, the preceding context has traditionally been treated differently in the literature. The context was partitioned differently, usually into the classes sibilants, stops, nasals, non-sibilant fricatives, and liquids, and the standard finding was that these contexts are ordered as follows in terms of descending [t, d]-deletion rates: sibilants > stops > nasals > non-sibilant fricatives > liquids. This particular order was most often claimed to be an effect of sonorancy – the less sonorous the preceding consonant the more likely [t, d] is to delete. However, there are several problems with this explanation.

Why would the sibilants that are more sonorous than the stops show higher deletion rates than the stops? This is no longer a mystery under the alternative view of the preceding context. The sibilants agree with [t, d] in place but differ from it in sonorancy. The stops agree with [t, d] in sonorancy, but differ from it in place. However, we have

seen above that co-occurrence of homorganic consonants are avoided more than co-occurrence of consonants that agree in sonorancy features.

The distinction between the sibilants and the non-sibilant fricatives are also no longer problematic. Although both of these groups have the same level of sonorancy, the sibilants also agree with [t, d] in place. Since the sibilants share more features with [t, d], [t, d] deletes more after the sibilants than after the non-sibilant fricatives.

The fact that the nasal context showed a higher [t, d]-deletion rate than the non-sibilant fricatives was also problematic. Nasals are certainly more sonorous than fricatives. An explanation can now also be offered for this. By far most of the nasals that precede [t, d] are [n]'s.[23] [n] shares two features with [t, d], while the non-sibilant fricatives share only one feature with [t, d]. We therefore expect higher [t, d]-deletion rates following [n] than following [f, v].

Overall, it seems that it is better to partition the preceding context following Guy and Boberg than the tradition in the literature. Unfortunately, since the literature generally partitions the preceding context differently than Guy and Boberg, it is not possible to determine beyond doubt whether the data from the rest of literature support the Guy and Boberg analysis. However, I have shown just above that the pattern claimed by Labov (1989:90) to be the standard (sibilants > stops > nasals > non-sibilant fricatives > liquids) is in general compatible with the Guy and Boberg analysis.[24]

---

[23] There are no monomorphemes that end in the sequence [-mt], [-md], [-ŋt] or [-ŋd]. And the number of bi-morphemes that end in these sequences are also much lower than the number of bi-morphemes that end in [-nd] or [-nt]. This is confirmed in the Guy and Boberg corpus. Their corpus contained 337 tokens with an [n] preceding a [t, d], and only 9 with an [m] or [ŋ] preceding [t, d].

[24] See §2.2.3.2 below for a more detailed discussion of other data from the literature.

2.2 The analysis

In this section I develop an analysis of the pattern in (25). The analysis will assume that constraints that are driving [t, d]-deletion are versions of the Obligatory Contour Principle (OCP) (Goldsmith, 1976, Leben, 1973, McCarthy, 1986, Yip, 1988). These constraints penalize contiguous identical features. The more features shared between [t, d] and a preceding consonant, the more severely it will be penalized by OCP-constraints. The rest of this section is structured as follows: In §2.2.1 I first discuss the constraints that are involved explaining the effect of the preceding context on [t, d]-deletion. Section §2.2.2 then contains the actual analysis within the rank-ordering model of EVAL. In section §2.2.3 I consider the typology of deletion patterns that are possible assuming the constraints formulated in §2.2.1. Finally, in §2.2.4 I consider two outstanding issues – the same issues that were also considered when I discussed the following context above (§1.2.4). First, I discuss the deletion of stops other than [t, d]. Secondly, I discuss the deletion of [t, d] preceded by a vowel.

2.2.1 The constraints

Since the phenomenon we are dealing with is one of deletion, the relevant faithfulness constraint is the anti-deletion constraint MAX. In order to explain why MAX is sometimes violated, we need markedness constraints that will be violated by the non-deletion candidates. These markedness constraints are specific instantiations of the OCP.

There is a long tradition in phonology of constraining the occurrence of contiguous identical elements. This idea was first implicated by Leben (1973) in order to explain the avoidance of contiguous identical tones. Goldsmith (1976) later termed the constraint the “Obligatory Contour Principle” to capture the fact that forms were required

to have tonal contours (i.e. no tonal plateaus). McCarthy (1986) and Yip (1988) showed that a similar principle also applies to segmental phonology – i.e. identical adjacent segments and identical features that are adjacent on some level of representation are also avoided.

Following Guy and Boberg (1997) I will assume that the markedness constraints that are responsible for the [t, d]-deletion pattern in (25) above are OCP-constraints against contiguous identical features, specifically against the three features that distinguish [t, d] from other consonants. The set relevant markedness constraints are defined in (26a) below. I am assuming the universal ranking in (26b) between these constraints. On this ranking, see below (26).

(26) a. **Markedness constraints**

*[+cor][+cor]: Do not allow two contiguous segments that are both specified as [+coronal].

*[-son][-son]: Do not allow two contiguous segments that are both specified as [-sonorant].

*[-cont][-cont]: Do not allow two contiguous segments that are both specified as [-continuant].

b. **Ranking**

||*[+cor][+cor] o *[-son][-son] o *[-cont][-cont]|| [25]

[25] The formulation of these OCP-constraints is different from the way in which Alderete (1996, 1997) and Itô and Mester (1997, 2003) formulate OCP-constraints. They define OCP-constraints as the local self-conjunction of markedness constraints (Smolensky, 1995) – i.e. if $*M$ is a markedness constraint, then $*M_{\delta}^{2}$ is an OCP-constraint that is violated whenever $*M$ is violated twice in domain δ. The existence of the OCP-constraint therefore depends on the existence of the un-conjoined markedness constraint $*M$.

The motivation for the ranking in (26b) comes from cross-linguistic patterns of consonantal co-occurrence restrictions. We know for Arabic (Frisch *et al*., 2004, Greenberg, 1950, Pierrehumbert, 1993), Hebrew (Greenberg, 1950), English, French and Latin (Berkley, 1994a, 1994b, 2000) that the co-occurrence of homorganic consonants is avoided more than the co-occurrence of consonants that agree in other features. Since *[+cor][+cor] penalizes co-occurrence of homorganic consonants, we can expect that it will outrank constraints on the co-occurrence of consonants agreeing in non-place features. We also know for Arabic that, within the class of coronals, those that agree in sonorancy are avoided more than those that agree in other features such as continuancy. Based on this, it can expected that *[-son][-son] outranks *[-cont][-cont].

In addition to the constraints in (26) that penalize contiguous segments that share one of the three features [+coronal, -sonorant, -continuant], there are also constraints that penalize contiguous segments that share two or three of these features. The motivation for these constraints again comes from cross-linguistic consonantal co-occurrence patterns. Pierrehumbert (Pierrehumbert, 1993), Frisch *et al.* (Frisch et al., 2004) and Berkley (Berkley, 1994a, 1994b, 2000) have all illustrated that consonantal co-occurrence restrictions are gradient – the more similar two consonants are, the less likely they are to

---

Gouskova (2003) has recently shown that assuming the existence of markedness constraints against "unmarked" features or segments results in incorrect typological predictions. The features [+coronal], [-sonorant] and [-continuant] are probably the most unmarked consonantal features. The unmarkedness of coronal place is generally accepted (de Lacy, 2002, 2003, Gnanadesikan, 1996, Jakobson, 1968, Paradis and Prunet, 1991, Prince, 1998). Assuming that the least marked consonants are as different from vowels as possible implies that stops are the least marked consonants. The features that select the stop consonants [-sonorant, -continuant] can therefore also be accepted as being unmarked features. If no markedness constraints against [+coronal] exists, then an OCP-constraint against two [+coronal] features can obviously not be formed via the local self-conjunction of a markedness constraint. The same is true of the features [-continuant] and [-sonorant]. There are therefore no *[+coronal], *[-continuant] or *[-sonorant] constraints that can be conjoined to form OCP-constraints. This is the why I do not define the constraints in (26) as locally self-conjoined markedness constraints.

co-occur in some domain. Consonants that share two of the features [+coronal, -sonorant, -continuant] with [t, d] are obviously more similar to [t, d] than consonants that share only one of these features with [t, d].[26] In order to capture these generalizations I assume the existence of the constraints in (27a) and (27b). I also assume that these constraints are ranked as shown in (27c) – a motivation of this ranking follows.

(27) **More markedness constraints**

a. **Sharing all three features**

***[+cor, -son, -cont][+cor, -son, -cont]**

Do not allow two contiguous segments that are both specified as [+coronal, -sonorant, -continuant].

b. **Sharing two of the features**

***[+cor, -son][+cor, -son]**

Do not allow two contiguous segments that are both specified as [+coronal, -sonorant].

***[+cor, -cont][+cor, -cont]**

Do not allow two contiguous segments that are both specified as [+coronal, -continuant].

***[-son, -cont][-son, -cont]**

Do not allow two contiguous segments that are both specified as [-sonorant, -continuant].

c. Ranking between the constraints in (26b)

||*[+cor, -son][+cor, -son] o *[+cor, -cont][+cor, -cont] o *[-son, -cont][-son, -cont]||

---

[26] Pierrehumbert (1993) calculates similarly between two consonants in terms of the number of shared features between the consonants. However, Frisch *et al*. (2004) show the number of natural classes shared by two consonants to be a better index of the similarity between the two consonants. Irrespective of which of these two measures of similarity is used, consonants that share two of the features [+coronal, -sonorant, -continuant] with [t, d] will be more similar to [t, d] than consonants that share only one of these features with [t, d].

Based on the fact that the co-occurrence of consonants is more restricted the more similar they are, I assume that all of the constraints in (27) rank higher than those in (26). On the same grounds, I also assume that the constraint in (27a) against sharing all three features ranks higher than the constraints in (27b) against sharing only two of the features.

The motivation for the ranking in (27c) comes again from the consonantal co-occurrence patterns in Arabic, English, French and Latin mentioned just above. In all of these languages co-occurrence of consonants that agree in place of articulation is avoided more than co-occurrence of consonants that do not agree in place of articulation. This motivates the ranking of the two constraints that include a reference to [+coronal] over the third constraint. At least for Arabic we know that within the set of coronals consonants that also agree in sonorancy are avoided more than those that do not agree in sonorancy. This is the motivation for the ranking between the first two constraints – since *[+cor, -son][+cor, -son] refers also to sonorancy, it outranks *[+cor, -cont][+cor, -cont].

In (28) I give a graphic representation of how all the markedness constraints that I assume are ranked.

(28) **The ranking between the markedness constraints in (26) and (27)**

Constraints	**Penalizes [t, d] preceded by …**
*[+cor, -son, -cont][+cor, -son, -cont]	[t, d]
|	
*[+cor, -son][+cor, -son]	[s, š, z, ž]
|	
*[+cor, -cont][+cor, -cont]	[n]
|	
*[-son, -cont][-son, -cont]	[k, g, p, b]
|	
*[+cor][+cor]	[l]
|	
*[-son][-son]	[f, v]
|	
*[-cont][-cont]	[m, ŋ]

2.2.2 Accounting for the observed patterns

Now that the relevant constraints have been identified, we can consider how these constraints interact to determine the variable [t, d]-deletion pattern shown in (23) and (25). As before, there are two aspects of the variation that need to be accounted for, namely the intra-contextual and the inter-contextual variation. The intra-contextual variation is discussed first in §2.2.2.1. Section §2.2.2.2 then deals with the inter-contextual variation.

2.2.2.1 Intra-contextual variation

The data in (23) show that, with the exception of the post [t, d] context, all contexts have variation between deletion and retention. In this section I first discuss the contexts that do show variation, and then return to the post [t, d] context.

There are two variants that are observed in the [t, d]-deletion phenomenon – the retention candidate that preserves [t, d], and the deletion candidate (indicated with the symbol ∅). In contexts where the deletion candidate is the more frequently observed variant, EVAL has to impose the harmonic ordering |∅ ™ t/d| on the candidate set. In contexts where the retention candidate is the more frequent variant, the opposite ordering has to be imposed on the candidate set, i.e. |t/d ™ ∅|. A glance back at (23) confirms that all the contexts that show variation are associated with deletion rates of lower than 50% – i.e. for all contexts we need the harmonic ordering |t/d ™ ∅|.[27]

---

[27] The deletion rate after sibilants is very close to 50%. However, in the discussion here I will assume that this context does actually show more retention than deletion. The fact that the deletion rate in this context is so close to chance indicates that this dialect of English is on the verge of changing from a dialect that prefers retention to a dialect that prefers deletion in this context. During this stage it might be that the second best candidate (the deletion candidate) is highly accessible – i.e. the deletion and the retention candidate do no differ much in their relative accessibility, so that individual speakers will show near chance deletion. It is also possible that individual speakers differ from each other. Some might have the ranking ||MAX o *[+cor, -son][+cor, -son]|| and therefore show more retention, while

In order to get the ordering |t/d ™ ∅| for these contexts, it is necessary that the deletion candidate always violate a higher ranked constraint than the retention candidates. Since the deletion candidate violates only MAX, we need MAX to outrank all of the markedness constraints that refer to these contexts. In order to assure that variation is observed in all contexts, it is also necessary that neither the deletion candidate nor the retention candidate be disfavored by any constraint ranked higher than the critical cut-off. This implies that the cut-off should be ranked higher than MAX and the markedness constraints. Lastly, we need to account for the fact that only two variants are observed, namely retention of [t, d] and deletion of [t, d]. In addition to deletion of [t, d], there are several other ways in which violation of the markedness constraints can be avoided. Consider an input such as /-Vlt/. The retention candidate for this input [-Vlt] violates *[+cor][+cor]. This violation can be avoided by deletion of the [t], i.e. [-Vl] (an actually observed variant), or by epenthesis of a vowel between the two consonants, i.e. [-Vl<u>V</u>t]. In order to prevent [-Vl<u>V</u>t] from surfacing as a variant, it is necessary that this candidate be disfavored by a constraint ranked higher than the cut-off. This candidate violates the anti-epenthesis constraint DEP. If we rank DEP higher than the cut-off, we can explain why [-Vl<u>V</u>t] is never observed as a variant output for /-Vlt/. In the discussion below, I will use this candidate as an example of a non-observed variant. All other non-observed variants are ruled out in the same manner – i.e. they all are disfavored by constraints ranked higher than the cut-off.

---

others have the opposite ranking ||*[+cor, -son][+cor, -son] o MAX|| and therefore show more deletion. Guy and Boberg (1997, see also Guy, 1994) report only the mean deletion rates for all the speakers together. It is therefore not possible to decide between these two possibilities.

Tableau (29) considers the evaluation of the relevant candidates for an input like /-Vlt/. In this tableau I include only one markedness constraint, namely *[+cor][+cor] – the constraint violated by the retention candidate [-Vlt]. Since all contexts with variation show the same variation pattern (more retention than deletion), this one tableau serves as an example for all these contexts. The variation pattern in other contexts can be explained in exactly the same way by replacing *[+cor][+cor] by the markedness constraint relevant to the particular context.

(29) **Variation for a /-Vlt/ input**

/-Vlt/	DEP	MAX	*[+cor][+cor]
1 -Vlt			*
2 -Vl __		*	
3 -Vl<u>V</u>t	*!		

Output of EVAL

-Vlt *[+cor][+cor]

-Vl __ MAX

Cut-off

-Vl<u>V</u>t DEP

Neither the retention candidate nor the deletion candidate violates any constraints ranked higher the cut-off. Both of these candidates are therefore predicted as possible outputs. Since EVAL rates the retention candidate better than the deletion candidate, the prediction is that the retention candidate will be the more frequently observed variant. The epenthetic candidate violates DEP which is ranked higher than the critical cut-off. Since there are candidates available that violate no constraints ranked higher than the cut-off, the epenthetic candidate will never be selected as an output variant.

Now we can consider the post [t, d] context. As explained when the data were discussed in §2.1, there are no forms in which a final [t, d] can actually be preceded by [t, d]. There are no monomorphemes in English with this sequence, and the past tense suffix used with verbs that end on [t, d] is /-əd/ rather than /-d/. However, because of "richness of the base" (Smolensky, 1996) we have to consider what the grammar would have done with inputs that contained one of the sequences /-td, -dd, -dt, -tt/. Since no forms like these exist in English, the assumption is that English does not tolerate sequences like these. Had such input sequences existed, they would then not be allowed to surface faithfully at all. Based on these considerations, I am assuming that a final [t, d] would have deleted 100% of the time from such sequences.

Unlike in the other contexts, the deletion candidate is therefore preferred over the retention candidate in the post [t, d] context. This means that for this context EVAL has to impose the ordering |∅ ™ t/d| on the candidate set. This is possible only if the retention candidate is disfavored by a constraint ranked higher than MAX. The markedness constraint that refers specifically to the post [t, d] context therefore outranks MAX, i.e. ||*[+cor, -son, -cont][+cor, -son, -cont] o MAX||. In order to assure that the deletion candidate is the only candidate that is accessed as output in this context, it is necessary that the retention candidate be disfavored by a constraint ranked higher than the cut-off, i.e. ||*[+cor, -son, - cont][+cor, -son, -cont] o Cut-off||. When the other contexts were discussed just above, I have argued that the Cut-off ranks higher than MAX. Combing these rankings therefore gives us ||*[+cor, -son, - cont][+cor, -son, -cont] o Cut-off o MAX||. Under this ranking a final [t, d] will always delete when it is preceded by [t, d]. This is shown in tableau (30).

(30) **Categorical deletion after [t, d]**

/-Vtt/	*[+cor, -son, - cont] [+cor, -son, -cont]	DEP[28]	MAX
3 -Vtt	*!		
1 -Vt __			*
2 -Vt<u>V</u>t		*!	

Output of EVAL

-Vt __ MAX

Cut-off

-Vt<u>V</u>t DEP

-Vtt *[+cor, -son, - cont][+cor, -son, -cont]

With the exception of the deletion candidate, all candidates are disfavored by a constraint ranked higher then cut-off. The deletion candidate violates only MAX, a constraint ranked lower than the cut-off. Since there is a candidate available that is not disfavored by any constraint ranked higher than the cut-off, the candidates that are disfavored by constraints ranked higher than the cut-off will not be accessed as outputs. The prediction is therefore that the deletion candidate will be the only observed output.

2.2.2.2 Inter-contextual variation

What remains to be accounted for is the difference between different contexts. Although all of the contexts show less than 50% deletion, the actual deletion rates in the different contexts are not the same. In (25) the contexts are listed in ascending order of [t, d]-deletion rate. In the rank-ordering model of EVAL such inter-contextual variation is

[28] We do not have evidence for the ranking between DEP and *[+cor, -son, - cont][+cor, -son, -cont]. Following the principle of ranking conservatism I rank the markedness constraint higher than the faithfulness constraint. On ranking conservatism, see the discussion in Chapter 4 §2.2.1.

accounted for by considering non-generated comparison sets – i.e. we compare the retention candidates from different input contexts with each other rather than different output candidates for a single input. This comparison is done in (31).

(31) **Comparison between retention candidates from different contexts**

Context			*[+cor, -son, -cont] [+cor, -son, -cont]	MAX	*[+cor, -son] [+cor, -son]	*[+cor, -cont] [+cor, -cont]	*[-son, -cont] [-son, -cont]	*[+cor][+cor]	*[-son][-son]	*[-cont][-cont]
[t, d]	7	-tt	*							
[s, š, z, ž]	6	-st			*					
[n]	5	-nt				*				
[k, g, p, b]	4	-kt					*			
[l]	3	-lt						*		
[f, v]	2	-ft							*	
[m, ŋ]	1	-mt								*

Output of EVAL

Context

[m, ŋ] -mt *[-cont][-cont]
|
[f, v] -ft *[-son][-son]
|
[l] -lt *[+cor][+cor]
|
[k, g, p, b] -kt *[-son, -cont][-son, -cont]
|
[n] -nt *[+cor, -cont][+cor, -cont]
|
[s, š, z, ž] -st *[+cor, -son][+cor, -son]
|
[t, d] -tt *[+cor, -son, -cont][+cor, -son, -cont]

The more marked the retention candidate is, the more can be gained by deletion of the [t, d]. The prediction is therefore that the deletion rates should steadily increase as we move down the rank-ordering that EVAL imposes on these candidates. In (25) the contexts were also arranged in terms of increasing [t, d]-deletion rate. Comparison between the ordering in (31) and the ordering in (25) will confirm that these two are the same. The predicted and observed relations between the contexts therefore agree. By allowing EVAL to compare the retention candidates from the different input contexts, we can account for the difference in deletion rates between the contexts.

This is a significant result. When I discussed the ranking between the different markedness constraints in §2.1 above, I showed that each of the rankings was motivated by consonantal co-occurrence restrictions in several languages. The constraints were therefore not ranked primarily based on the [t, d]-deletion patterns. The fact that this particular ranking correctly accounts for the inter-contextual variation can therefore be taken as strong evidence in favor of the analysis presented here.

2.2.3 Factorial typology – what are the possible dialects?

I have argued above that we need seven markedness constraints, one faithfulness constraint and the critical cut-off to account for the [t, d]-deletion pattern discussed in §2.2.1 and §2.2.2. We have to consider the factorial typology that follows from this – i.e. what are all the possible rankings between these constraints and the cut-off and what are the deletion patterns that would result from each of these rankings? These and only these patterns are predicted as possible.

Since no dialect of English ever allows a word to end on two coronal stops, I will assume that the post [t, d] context will always be associated with categorical deletion, and

will not consider this context here. I will therefore consider only six of the markedness constraints here. Although there are six markedness constraints, I have argued that the ranking between these constraints are universally fixed as in (28). All that we have to consider is how the critical cut-off and MAX can be ranked relative to this fixed ranking. There are seven positions on the ranking in (28) where MAX can be ranked. In each of these seven rankings there are eight positions where the critical cut-off can be located. This means that there is a total of 56 possible rankings to consider. I will not discuss each of these 56 grammars. However, I will point out all the deletion patterns that are predicted as possible (that result from at least one of these rankings), and also some of the deletion patterns that are predicted as impossible (that would not result from any of the possible rankings).

The rest of this section is structured as follows: I begin by considering what the conditions are that must be satisfied for any variation to be observed (§2.2.3.1). After that has been established, I will first discuss the deletion patterns that are predicted as possible (§2.2.3.2), and then the patterns that are predicted as impossible (§2.2.3.2).

2.2.3.1 Conditions for variation

Variation is only possible if there is more than one candidate that is not disfavored by any constraint ranked higher than the critical cut-off. In the [t, d]-deletion phenomenon the observed variants are the retention and the deletion candidates. For some context to show variation, it is therefore necessary that neither the deletion candidate nor the retention candidate for that context be disfavored by a constraint ranked higher the cut-off. The deletion candidate will always violate MAX. The first requirement is therefore that MAX should rank below the cut-off. The retention candidate in each context will violate one or

more of the six markedness constraints assumed here.[29] In order for the retention candidate to be one of the observed variants, it is also necessary that all the markedness constraints that it violates rank lower than the cut-off. These requirements are stated in a general form in (32).

(32) **Necessary conditions for variation in each context**

a. **General condition**

||Cut-off ∘ MAX||

b. **Context specific conditions**

Let $\mathbb{M}$ stand for the set of markedness constraints violated by the retention candidate in some context. Then variation in this context can be observed only if:

For all $M \in \mathbb{M}$: ||Cut-off ∘ M||

2.2.3.2 Possible deletion patterns

I list the deletion patterns that are predicted as possible in (33). I do not show here why these are all and only the possible patterns. This discussion can be found in the Appendix at the end of this chapter. In the tables in (33) "D" stands for "categorical deletion", "R" for "categorical retention", "D > R" for "more deletion than retention" and "R > D" for "more retention than deletion". In order to make it easier to identify patterns, I also shade all cells in which more deletion than retention is observed – i.e. both "D" and "D > R"

[29] Whenever the retention candidate violates one of the constraints that refer to more than one feature, it will also violate the constraints that refer to each of these features individually.

cells. The different contexts that can precede a final [t, d] are listed in the top row in the order of decreasing markedness.

(33) **Possible deletion patterns**

a. **No-variation**

[s, š, z, ž]	**[n]**	**[k, g, p, b]**	**[l]**	**[f, v]**	**[m, ŋ]**
R	R	R	R	R	R
D	R	R	R	R	R
D	D	R	R	R	R
D	D	D	R	R	R
D	D	D	D	R	R
D	D	D	D	D	R
D	D	D	D	D	D

b. **Variation in all contexts**

[s, š, z, ž]	**[n]**	**[k, g, p, b]**	**[l]**	**[f, v]**	**[m, ŋ]**
R > D	R > D	R > D	R > D	R > D	R > D
D > R	R > D	R > D	R > D	R > D	R > D
D > R	D > R	R > D	R > D	R > D	R > D
D > R	D > R	D > R	R > D	R > D	R > D
D > R	D > R	D > R	D > R	R > D	R > D
D > R	D > R	D > R	D > R	D > R	R > D
D > R	D > R	D > R	D > R	D > R	D > R

c. **Variation only in some contexts**

[s, š, z, ž]	**[n]**	**[k, g, p, b]**	**[l]**	**[f, v]**	**[m, ŋ]**
D	R > D	R > D	R > D	R > D	R > D
D	D > R	R > D	R > D	R > D	R > D
D	D > R	D > R	R > D	R > D	R > D
D	D > R	D > R	D > R	R > D	R > D
D	D > R	D > R	D > R	D > R	R > D

((33c) continued)

[s, š, z, ž]	[n]	[k, g, p, b]	[l]	[f, v]	[m, ŋ]
D	D > R	D > R	D > R	D > R	D > R
D	D	R > D	R > D	R > D	R > D
D	D	D > R	R > D	R > D	R > D
D	D	D > R	D > R	R > D	R > D
D	D	D > R	D > R	D > R	R > D
D	D	D > R	D > R	D > R	D > R
D	D	D	R > D	R > D	R > D
D	D	D	D > R	R > D	R > D
D	D	D	D > R	D > R	R > D
D	D	D	D > R	D > R	D > R
D	D	D	D	R > D	R > D
D	D	D	D	D > R	R > D
D	D	D	D	D > R	D > R
D	D	D	D	D	R > D
D	D	D	D	D	D > R

The most striking characteristic of these tables is that if a cell is shaded, then all cells to its left are also shaded. This means that if some context is associated with more deletion than retention, then so are all contexts in which retention would be more marked. This is a very reasonable prediction. If the markedness of the retention candidate in some context is severe enough that deletion is preferred over retention for that context, then the same should be true of contexts in which the retention candidate is even more marked.

A prediction that is not shown in the tables in (33) is that the deletion rate should steadily fall across the contexts from left to right – that is, the highest deletion rate is expected for the most marked sequence, and lower deletion rates for less marked sequences. This is clearly shown in (31) above. In tableau (31) I compared the retention

candidates from the different contexts for the Guy and Boberg (1997) corpus. Although the comparison there was done specifically for their corpus, it can stand in for any possible grammar. Since the ranking between the markedness constraints is fixed, the comparison between the retention candidates will always result in the same rank-ordering between the retention candidates, irrespective of where MAX and the critical cut-off rank amongst the markedness constraints.

The only one of the patterns in (33) that we know for certain to exist is that represented by the first line of table (33b) – variation in all contexts with more retention than deletion in all contexts. This is the pattern that is observed in the speech of the three Philadelphians that Guy and Boberg (1997, see also Guy, 1994) report on and that was discussed in detail in §2.1 and §2.2.2 above. There is no other dialect of English for which we have adequate information to determine whether it exemplifies one of the other patterns. The reason for this is, of course, that the other studies of [t, d]-deletion partitions the preceding context differently – see the discussion in §2.1 about this. However, it is at least possible to determine for some dialects discussed in the literature whether they are likely to fit into one of these patterns. The tables in (34) are a representative sample of the data from the literature on the influence of the preceding context. The classes are listed in order of descending deletion rate from left to right.[30]

[30] Labov *et al.* (1968) report on the English of African American and Puerto Rican speakers in New York City. Unfortunately they do not report the actual deletion rates associated with different preceding contexts. However, they do state that [t, d] deletes most often after [s], and very infrequently after [l] (which they ascribe to the fact that post-vocalic [l] is most often "non-consonantal") (p. 129).

(34) **Influence of the preceding context on the [t, d]-deletion rate**

(a) Neu data (1980:49)

Class	**[s, š, z, ž]**	**[p, b, k, g]**	**[n, m, ŋ]**	**[l]**	**[f, v]**
n	306	123	693	117	70
% deleted	37	31	30	5	3

(b) Jamaican English (Kingston) (Patrick, 1991:178)

Class	**[s, š, z, ž]**	**[p, b, k, g]**	**[f, v]**	**[n, m, ŋ]**	**[l]**
n	462	162	73	907	168
% deleted	85	80	75	67	58

(c) Tejano English (San Antonio) (Bayley, 1997:310)

Class	**[s, š, z, ž]**	**[m, n, ŋ]**	**[p, b, k, g]**	**[l]**	**[f, v]**
n	1,288	1,073	334	375	206
% deleted	72	40	33	21	16

(d) Trinidadian English (Kang, 1994:157)

Class	**[s, š, z, ž, k, g, p, b, n, m, ŋ]**	**[f, v, l]**
n	51	30
% deleted	45	30

(e) African American English (Washington, DC) (Fasold, 1972:70)

Class	**[n, m, ŋ, l]**	**[s, š, z, ž, f, v]**	**[p, b, k, g]**
n	147	112	123
% deleted	63	49	37

(f) Chicano English (Los Angeles) (Santa Ana, 1991:92)

Class	**[m, n, ŋ]**	**[s, š, z, ž]**	**[l]**	**[f, v]**	**[r]**	**[p, b, k, g]**
n	1,779	777	354	83	475	21
% deleted	67	60	34	16	14	5

With the exception of the Chicano English in (34f), all of these dialects are compatible with the predictions of the analysis developed above. Consider first the Neu

data from (34a). As predicted by the analysis, post-sibilant context shows the highest deletion rate. Since this context still shows less than 50% deletion, it means that these data represent a dialect that fall in the first line of (33b) – i.e. more retention than deletion everywhere. The prediction of the analysis here is that [n] should show more deletion than the stops, and the stops again more than [m, ŋ]. Since Neu groups all three nasals together it is not possible to determine whether this particular prediction is confirmed by her data. She reports near equal deletion after the nasals and after the stops. It is possible that a higher deletion rate after [n] and the lower deletion rate of [m, ŋ] average out to a deletion rate that is about equal to the deletion rate after the stops. My analysis predicts a higher deletion rate after [l] than after the non-sibilant fricatives. This is confirmed in Neu's data. Since Neu groups [m, ŋ] together with [n], it is not possible to determine what the deletion rate is that is associated with [m, ŋ] alone. We can therefore not determine whether post [m, ŋ] context shows a lower deletion rate than post [f, v] context.

The Jamaican English data in (34b) would fall in the last of the possible patterns in (33b) – with more deletion than retention everywhere. As predicted, it has the highest deletion rate after sibilants. Next we would expect [n]. However, since [m, ŋ] is lumped together with [n], it is very likely the case that the low deletion rate after [m, ŋ] pulls down the mean rate for nasals. The most unexpected result is that [l] shows less deletion than [f, v]. However, it is possible that the [l] of Jamaican English is often vocalized, in which case we would expect a low deletion rate after [l] – see for instance Labov *et al*. (1968:129) who use this as explanation for the low deletion rate after [l] in the AAE of Harlem.

Tejano English (34c) represents a dialect from the second pattern in (33b) – more deletion than retention in the most marked post-sibilant context, more retention than deletion everywhere else.

For Trinidadian English (34d) it is not possible to decide which pattern it would represent – both since the preceding context is partitioned into only two groups, and because the number of tokens in this corpus is very small.

Washington, DC AAE (34e) is most likely representative of the pattern in the third line of (33b) – more deletion than retention after sibilants and [n], but more retention than deletion elsewhere. Fasold reports a 63% deletion rate after [n, m, ŋ, l]. Since [l] and [m, ŋ] are expected to have low deletion rates, it means that the actual deletion rate after [n] is probably even higher. For the class of fricatives [s, š, z, ž, f, v] Fasold reports a 49% deletion rate. The actual rate after sibilants is almost certainly higher, and is pulled down by a relatively low deletion rate after [f, v]. Fasold reports a rate below 50% for the stops, and we would therefore expect [l], [f, v] and [m, ŋ] also to show less than 50% deletion – since retention in all of these contexts are less marked than retention after a stop, and these contexts should therefore show even less deletion than the post-stop context.

The data for Chicano English in (34f) are problematic. The deletion pattern in these data goes against the predictions of the analysis above in several respects. First, the nasals show a higher deletion rate than the sibilants. Secondly the stops show a lower deletion rate than [l], [f, v] and [r]. These data do not only counter predictions of the analysis that I developed above, but also the statement of Labov about how the preceding context in general influences [t, d]-deletion. Labov (1989:90) summarizes the results of studies on [t, d]-deletion, and notes that preceding consonants can be ordered as follows

in terms of decreasing deletion rates: sibilants > stops > nasals > non-sibilant fricatives > liquids. Santa Ana (1996:72) also acknowledges that his data on Chicano English counter all other data reported in the literature. The Chicano data therefore represent a truly unexpected pattern.[31]

To the best of my knowledge no dialect has been reported with one of the patterns from (33a) (variation in no context) or from (33c) (variation only in some contexts). Even so, all of these patterns seem reasonably possible – they all show at least as much deletion in the more marked than in the less marked contexts.

2.2.3.3 Impossible deletion patterns

We also have to consider deletion patterns that are predicted as impossible. These are probably even more instructive. If a deletion pattern that is predicted as impossible does actually exist, it would count as rather strong evidence against the analysis.

A group of deletion patterns that are predicted as impossible are patterns in which a more marked sequence is associated with lower deletion rates than a less marked sequence. I list a few of these patterns as examples in (35).

(35) **Impossible patterns: More retention in more marked context than in less marked context**

[s, š, z, ž]	[n]	[k, g, p, b]	[l]	[f, v]	[m, ŋ]
R	R	R	D	D	D
R	R	R	R > D	R > D	R >D
D > R	D > R	D > R	D	D	D
R > D	R > D	R > D	D	D	D
R > D	R > D	R > D	D > R	D > R	D > R

[31] See the next section for more discussion of the Chicano deletion pattern.

The patterns in (35) represent only a portion of the large number of patterns like these that are predicted as impossible. To make the discussion more concrete, consider the deletion rates after [n] and after [m, ŋ] in the patterns in (35). In all of these examples [t, d] would delete less after [n] than after [m, ŋ]. All of these patterns are predicted as impossible under the analysis developed above. This prediction follows straightforwardly from the universal ranking between the markedness constraints in (28). Because *[+cor, -cont][+cor, -cont] universally outranks *[-cont][-cont], retention after an [n] will always be more marked than retention after [m, ŋ]. With the exception of the deletion pattern of Chicano English (see (34f) above), I know of no dialect of English that counters this prediction.

The Chicano English data do counter this prediction, and therefore represent a potential problem for the analysis that I developed above. For instance, in Chicano English [t, d] deletes more after nasals than after sibilants. However, we should remember that factorial typology predictions assume that the constraints being considered are the only relevant constraints. It is possible that there are other constraints that could also have an influence on [t, d]-deletion. Santa Ana (1996) suggests that the Chicano English pattern is governed not by OCP-constraints, but by constraints on syllabic well-formedness. Following Clements (1988), Santa Ana claims that a high sonority coda is more well-formed than a low sonority coda. Deletion of a final [t, d] after a segment that is high in sonority will therefore result in a syllable with a highly desirable coda. However, deletion after a consonant that is low in sonority will result in a syllable with a coda that is less desirable. This could explain why [t, d] deletes more after nasals than sibilants. Deletion after a nasal results in a syllable with a highly desirable nasal coda.

Deletion after a sibilant results in a syllable with a less desirable sibilant coda. It therefore seems to be the case that in Chicano English the constraints on the well-formedness of codas outrank the OCP-constraints that were formulated above in §2.1.1. As a result, these syllabic well-formedness constraints would take precedence in deciding the deletion pattern.[32] Even if this could account for the Chicano pattern, it would still be hard to explain why Chicano English differs from all other dialects. Currently I can offer no explanation for this.

There is another group of patterns that are predicted as impossible, namely patterns in which some context shows variable deletion and some less marked context shows categorical retention. I list a few of these patterns as examples in (36).

(36) **Impossible patterns: Variable deletion in some context, and categorical retention in a less marked context**

[s, š, z, ž]	[n]	[k, g, p, b]	[l]	[f, v]	[m, ŋ]
R > D	R > D	R > D	R	R	R
D > R	D > R	D > R	R	R	R
D > R	D > R	R > D	R > D	R	R

All of the examples in (36) show variable deletion at the more marked end of the spectrum, and categorical retention at the less marked end. To make the example more concrete, consider only the post [n] and post [f, v] contexts. In the patterns in (36) there is

[32] Let OCP stand for the OCP-constraints defined in §2.2.1 above, and SWF stand for the syllabic well-formedness constraints that determine the Chicano deletion patterns. In particular let SWF_{Nasal} stand for the constraint on nasal codas, SWF_{Sib} for the constraint on sibilants, and SWF_{Rest} for all the other coda constraints grouped together. Chicano English would then need the following ranking: ||Cut-off o SWF_{Nasal} o SWF_{Sib} o MAX o SWF_{Rest} o OCP||. Since the cut-off ranks above both MAX and the SWF constraints, it follows that there will be variation between deletion and retention. Since SWF_{Nasal} and SWF_{Sib} both outrank MAX, EVAL will order the candidates in these contexts as follows: |∅ ™ t/d| – i.e. we expect more deletion than retention here. Since MAX outranks the rest of the SWF constraints, EVAL will impose the opposite order on the candidate set for these contexts, i.e. |t/d ™ ∅|. For these contexts we would then expect more retention than deletion. The OCP-constraints have to rank relatively low – so that they cannot override the effects of the SWF constraints.

variable deletion after [n] and categorical retention after [f, v]. The impossibility of these patterns is a somewhat unexpected prediction. Retention after [f, v] violates *[-son][-son] and retention after [n] violates *[+cor, -cont][+cor, -cont]. Because of the universal ranking ||*[+cor, -cont][+cor, -cont] o *[-son][-son]|| (see (28)), retention after [n] will always be more marked than retention after [f, v]. From this we expect that [t, d] will always delete more after [n] than after [f, v]. Why is it then not possible for [t, d] to be categorically retained after [f, v] and to delete variably after [n]?

To see how this prediction follows from the analysis developed above, consider what the conditions are that have to be met for the [n] context to show variable deletion. Both of the variation conditions from (32) have to be met for this context, i.e. we need the ranking ||Cut-off o {MAX, *[+cor, -cont][+cor, -cont]}||. However, the universal ranking ||*[+cor, -cont][+cor, -cont] o *[-son][-son]|| (see (28)) entails through transitivity of constraint ranking that we also have ||Cut-off o {MAX, *[-son][-son]}||. Both variation conditions are therefore also met for the post [f, v] context, so that it is not possible to have categorical retention in this context.[33] In general then, if some context shows variation, then so will all less marked contexts. Although this prediction of my analysis seems potentially problematic, to the best of my knowledge no dialect has been reported that shows any of these patterns. For now, I therefore only acknowledge this as a falsifiable prediction of the analysis.

[33] Categorical deletion is also excluded in the same manner. However, this is not a problem. If post [n] context had variable deletion and post [f, v] context had categorical deletion, then the less marked post [f, v] context would have had more deletion than the more marked post [n] context. See the discussion earlier in this section about cases like these.

2.2.4 Two outstanding questions

When I discussed the influence of the following context, I showed that the other stops of English [p, b, k, g] also delete, and that a [t, d] deletes also when it is preceded by a vowel (§1.2.4). I will therefore not again show that this is true. However, I will briefly consider the implication of these facts for the analysis of the influence of the preceding context developed above.

2.2.4.1 What about [p, b] and [k, g]?

I have already shown in §1.2.4.1 that the labial and velar stops of English are also subject to variable deletion when they occur as final member of a word-final consonant cluster. As I also mentioned then, there are no data available on the deletion patterns associated with the non-coronal stops. The discussion in this section is therefore necessarily speculative.

The most straightforward assumption would be that the deletion of the non-coronal stops is influenced by the same factors that influence the deletion of the coronal stops. I argued above that the constraints that drive the deletion of the coronal stops are constraints of the OCP-family – constraints against contiguous segments that share certain features. In order to determine what the constraints are that influence the deletion of the labial and the velar stops, it is first necessary to determine what the features are that uniquely identify these two classes of stops. This is shown in (37).

(37) **Distinctive features of the non-coronal stops**

[p, b] = [+labial, -sonorant, -continuant]

[k, g] = [+velar, -sonorant, -continuant]

From these features we can now form OCP-constraints similar to the constraints that were formulated above for the coronals. The relevant constraints are listed in (38).

(38) **Markedness constraints that influence the deletion of non-coronal stops**

a. **Applying to both labials and velars**

*[-son][-son]:	Do not allow two contiguous segments that are both specified as [-sonorant].
*[-cont][-cont]:	Do not allow two contiguous segments that are both specified as [-continuant].
*[-son, -cont][-son, -cont]:	Do not allow two contiguous segments that are both specified as [-sonorant, -continuant].

b. **Specific to labials**

*[+lab][+lab]:	Do not allow two contiguous segments that are both specified as [+labial].
*[+lab, -son][+lab, -son]:	Do not allow two contiguous segments that are both specified as [+labial, -sonorant].
*[+lab, -cont][+lab, -cont]:	Do not allow two contiguous segments that are both specified as [+labial, -continuant].

c. **Specific to velars**

*[+vel][+vel]:	Do not allow two contiguous segments that are both specified as [+velar].
*[+vel, -son][+vel, -son]:	Do not allow two contiguous segments that are both specified as [+velar, -sonorant].
*[+vel, -cont][+vel, -cont]:	Do not allow two contiguous segments that are both specified as [+velar, -continuant].

Velar and labial stops share the features [-sonorant, -continuant] with coronals, so that the constraints in (38a) apply to stops at all three places of articulation. The constraints in (38b) and (38c) refer specifically to place and therefore apply only to some stops. In the discussion on coronals, I argued that agreement in place is avoided most, then in sonorancy, then in continuancy. Based on this I list I list the rankings for labials and velars in (39). Next to each constraint I list the contexts in which it applies. Labial and velar stops occur only after sibilants and homorganic nasals (Fudge, 1969:268). I indicate contexts in which they do not occur by striking out the segments representing these contexts.[34]

(39) **Ranking of the constraints for labials and velars**

Labials		**Velars**	
Constraints	**Penalizes [p, b] preceded by …**	**Constraints**	**Penalizes [k, g] preceded by …**
*[+lab,-son][+lab,-son]	[~~f, v~~]	*[+vel,-son][+vel,-son]	_[35]
|		|	
*[+cor,-cnt][+cor,-cnt]	[m]	*[+vel,-cnt][+vel,-cnt]	[ŋ]
|		|	
*[-son,-cnt][-son,-cnt]	[~~t, d, k, g~~]	*[-son,-cnt][-son,-cnt]	[~~t, d, p, b~~]
|		|	
*[+lab][+lab]	_[36]	*[+vel][+vel]	_[37]
|		|	
*[-son][-son]	[s, ~~š, z, ž~~]	*[-son][-son]	[s, ~~š, z, ž, f, v~~]
|		|	
*[-cont][-cont]	[~~n, ŋ~~]	*[-cont][-cont]	[~~n, m~~]

[34] An example of a non-occurring sequence with a labial stop is [-fp]. Since this sequence is not allowed in English, there are no monomorphemes with this sequence. Since English does not have a labial stop suffix, /-p/ or /-b/, no bi-morphemes with this sequence is possible either. All of these non-occurring sequences must be ruled out by different markedness constraints that are ranked higher than the cut-off.

[35] This cell would have contained segments with the feature combination [+velar, -sonorant, +continuant], i.e. velar fricatives. English does not have any velar fricatives, so that this cell is empty.

[36] This cell would have contained [+labial, +sonorant, +continuant] segments, i.e. a labial counterpart of [l]. Since the English consonantal inventory does not contain any such sounds, this cell is empty.

[37] This cell would have contained segments with the features combination [+velar, +sonorant, +continuant], i.e. a velar counterpart of [l]. Since the English consonantal inventory does not contain any such sounds, this cell is empty.

From these rankings it is easy to predict which contexts would be associated with higher deletion rates. The higher ranked the markedness constraints that would be violated by a retention candidate, the more marked retention in that context would be, and the more deletion we would expect.[38] In (40) I list for each of the places of articulation the different preceding contexts in descending order of predicted deletion rate. The order for labials and velars comes from (39), while the order for the coronals is from (31).

(40) **Predicted order in deletion rate for different places of articulation**

[t, d] after …	**[p, b] after …**	**[k, g] after …**
[s, š, z, ž]	[~~f, v~~]	–
[n]	[m]	[ŋ]
[k, g, p, b]	[~~t, d, k, g~~]	[~~t, d, p, b~~]
[l]	–	–
[f, v]	[s, ~~š, z, ž~~]	[s, ~~š, z, ž, f, v~~]
[m, ŋ]	[~~n, ŋ~~]	[~~n, m~~]

More deletion

The prediction is therefore that the deletion pattern after velar and labial stops would be quite different from the pattern after coronal stops. For coronals we expect the highest deletion rate after sibilants and [n]. For labials and velars these contexts are predicted to have the lowest deletion rates. Of course, since labial and velar stops do not occur after heterorganic nasals, the prediction about [n] is not testable. However, the

[38] Refer to the inter-context comparison for coronals in (31) above for an illustration of how this prediction can be derived from an OT tableau.

labial and velar stops do occur after [s]. In principle this prediction could be tested – does [t, d] delete much more frequently after [s] than either [p, b] or [k, g] does?

Another prediction is that [t, d] should delete relatively infrequently after [m], while [p, b] should delete much more after [m]. Similarly, [t, d] should delete infrequently after [ŋ] while [k, g] should delete much more after [ŋ]. Since labial stops can be preceded by [m] and velar stops by [ŋ], these two predictions are also in principle testable.

The analysis developed above for [t, d]-deletion makes interesting testable predictions about the deletion of velar and labial stops. This suggests worthwhile ways in which to expand the research on variable stop deletion in English.

2.2.4.2 What about [t, d] preceded by a vowel?

In §1.2.4.2 I have already shown that word-final [t, d] also deletes when it is preceded by a vowel. Since vowels do not share any of the features [+coronal, -sonorant, -continuant] with [t, d], a [t, d] preceded by a vowel will be unmarked relative to a [t, d] preceded by any consonant. The prediction is therefore that this context will show the lowest deletion rate of all contexts. This prediction is confirmed in Guy and Boberg's (1997) corpus. They do not list the actual deletion rate of [t, d] after vowels. However, they do state that this context shows "nearly categorical retention" (p. 155).

3. Considering alternatives

There are three alternative proposals for dealing with variable data in OT, namely crucially unranked constraint grammars (Anttila, 1997, Anttila and Cho, 1998, Anttila and Revithiadou, 2000, Anttila and Fong, 2002), floating constraint grammars (Nagy and Reynolds, 1997, Reynolds, 1994, Reynolds and Sheffer, 1994), and stochastic grammars (Boersma, 1998, Boersma and Hayes, 2001, Zubritskaya, 1997). In this section I will compare these alternatives to the rank-ordering model of EVAL, showing that they face certain conceptual and practical problems in explaining the kind of variable data that were discussed in this and the preceding chapters. This section therefore serves as a part not only of this chapter, but also of Chapter 3 on Latvian and of Chapter 4 on Portuguese. Before discussing the problems associated with these alternative models, I will first give a very brief review of how variation is accounted for in each of these models.

Crucially unranked constraints. Anttila (Anttila, 1997) proposed an extension to the classic OT grammar that could account for variable phenomena. He assumes that the constraint hierarchy of some language can contain a set of crucially unranked constraints. On every evaluation occasion one total ordering of the constraints is chosen from among the different total orderings possible between the crucially unranked constraints. Each of the possible total orderings is equally likely to be chosen on any evaluation occasion. Variation arises if different rankings between these crucially unranked constraints select different candidates as optimal. This theory not only explains how variation can arise, but it also makes predictions about the absolute frequency with which each variant for some input will be observed. Assume that there are *n* crucially unranked constraints. There are then *n*! possible rankings between these *n* constraints. Suppose that there are two variants,

and that m of the $n!$ rankings select $variant_1$ as optimal and $(n! - m)$ rankings select $variant_2$ as optimal. The prediction is then that $variant_1$ will be observed $m/n!$ of the time, and that $variant_2$ will be observed $(n! - m)/n!$ of the time.

Floating constraints. Independently from Anttila, Reynolds proposed a very similar extension to classic OT (Reynolds, 1994). Rather than having a set of crucially unranked constraints, Reynolds assumes that there are "floating constraints". A floating constraint is a constraint that is crucially unranked relative to a span of the (ranked) constraint hierarchy (this span is known as the "floating range" of the floating constraint). On every evaluation occasion the floating constraint is ranked in one specific location along its floating range. Variation arises if different docking sites for the floating constraint results in different candidates being selected as output. As in the Anttila model, the Reynolds model also makes predictions about the absolute frequency with which different variants will be observed. If there is one floating constraint with a floating range that is n constraints long, then there are $(n + 1)$ docking sites for the floating constraint along its floating range. Suppose that there are two variants, and that m of the docking sites for the floating constraint result in $variant_1$ being selected as optimal and $(n + 1 - m)$ of the docking sites result in $variant_2$ being selected as optimal. The prediction is then that $variant_1$ will be observed $m/(n + 1)$ of the time, and that $variant_2$ will be observed $(n + 1 - m)/(n + 1)$ of the time.

Stochastic OT grammars. In a stochastic OT grammar (Boersma, 1998, Boersma and Hayes, 2001) constraints are ranked along a continuous ranking scale. Every constraint has a basic ranking value along this scale. The actual point where a constraint is ranked along the continuous ranking scale is not equivalent to its basic ranking value.

A stochastic OT grammar includes a noise component – on every evaluation occasion a (positive or negative) random value is added to the basic ranking value of every constraint. This noise has a normal distribution with zero as its mean and some arbitrarily chosen standard deviation that is set at the same value for all constraints. Since the actual ranking value for some constraint is the result of adding this noise to the basic ranking of the constraint, the actual ranking values of a constraint will also be normally distributed around its basic ranking value. Since the ranking of a constraint is not fixed on the continuous ranking scale two constraints C_1 and C_2 can be ranked $||C_1 \circ C_2||$ on one evaluation occasion but $||C_2 \circ C_1||$ on the next evaluation occasion. Variation arises when one of these rankings selects one candidate as optimal, while the other ranking selects another candidate as optimal. The likelihood of either $||C_1 \circ C_2||$ or $||C_2 \circ C_1||$ being observed can be controlled very precisely by varying the distance between the basic ranking values of C_1 and C_2. In this way a stochastic OT grammar also makes predictions about the absolute frequencies with which different variants are observed. In fact, by controlling the distance between the basic ranking values of C_1 and C_2, any frequency distribution between the two rankings can be modeled.

There are two obvious differences between these alternatives and the rank-ordering model of EVAL that I am proposing. The first is a more conceptual difference about the locus of variation – is variation situated in the grammar or outside of the grammar? The second difference has to do with the question of whether grammar should account for absolute frequencies or only for relative frequencies.

A basic conceptual difference between the rank-ordering model of EVAL and the other models is where variation is seated. In OT the grammar of some language can be

equated with a ranking of the constraint set. In the other three models variation is then seated directly in the grammar. Variation arises because the constraints are ranked differently at different evaluation occasions. Consider a language where one input can have two outputs, $variant_1$ and $variant_2$. On one evaluation occasion the grammar of this language will rate these candidates as follows: |$variant_1$ ™ $variant_2$|. In this situation $variant_1$ will be the observed output. On the next evaluation occasion, the ranking between the constraints can be different and the grammar can then rate the candidates differently, i.e. |$variant_2$ ™ $variant_1$|. In this situation $variant_2$ will be the output. Variation arises because the grammar is different and imposes a different information structure on the candidate set at different evaluation occasions. The source of variation is therefore situated directly in the grammar.

The rank-ordering model of EVAL is very different in this regard. For any language there is only one ranking between the constraints and therefore only one grammar. The grammar will impose exactly the same information structure on the candidate set at every evaluation occasion. If there are two variants, $variant_1$ and $variant_2$, and $variant_1$ is observed more frequently than $variant_2$, then the grammar will on every evaluation occasion impose the rank-ordering |$variant_1$ ™ $variant_2$| on the candidate set. Variation arises in the way in which the language user interacts with this invariant output of the grammar. The locus of variation is then situated outside of grammar proper. Grammar specifies the limits within which variation will be observed – it specifies which candidates are possible variants (by the critical cut-off) and it specifies the relative frequency between the variants. But the actual output of the grammar is invariant.

A second difference between the alternatives and the rank-ordering model of EVAL is that the alternatives attempt to account for the absolute frequency with which different variants are observed, while the rank-ordering model of EVAL only accounts for the relative frequencies of the different variants. If there are two variants and they are ranked as |*variant*$_1$ ™ *variant*$_2$| by EVAL, then the prediction is that *variant*$_1$ will be observed more frequently than *variant*$_2$. However, no prediction is made about how much more frequent *variant*$_1$ will be. The other three models account for absolute frequencies, while the rank-ordering model of EVAL accounts only for relative frequencies. It seems that these models account for more aspects of the variable data and that they are therefore better. This is only an apparent advantage of these models over the rank-ordering model of EVAL. In the rest of this section I will show that these models face certain conceptual and practical problems precisely because they make predictions about absolute frequencies. The rank-ordering model of EVAL avoids all of these problems.

3.1 Conceptual problems

I will discuss two conceptual problems. The first I will call the "which-frequency" problem, and the second the "grammar-alone" problem.

The data on variation that are reported in the literature usually represent the average pattern associated with some speech community. However, it is highly unlikely that all (or even any) of the individuals that make up that speech community will have exactly the same variation pattern as the community average. As an example, consider again the Latvian example discussed in Chapter 3. In Latvian, unstressed vowels are variably deleted from final unstressed syllables. The data that we have on this process

were collected by Karinš (1995a) from eight individual speakers of the Riga dialect of Latvian. Karinš found that his eight subjects deleted the vowel on average 86% of the time. However, there was considerable variation between the eight subjects. The standard deviation is his data was 9.9%, and the subject with the lowest deletion rate deleted only 67% of the time while the subject with the highest deletion rate deleted 97% of the time.

In this dataset we therefore have the average deletion rate of 86%, but we also have the deletion rates of the eight individual subjects that differ from each other and from the community average. Of all these different absolute deletion rates, which one should be modeled by the grammar? One possibility is to model the average of 86%. However, it is debatable how real this average actually is. The grammar that we are then modeling is not the actual grammar of any individual member of the community. As an alternative we can model the grammars of the different individual members separately. But this tips the scale too far in the other direction. We then lose the concept that the members of the speech community do share a grammar.

What is it that the eight individuals studied by Karinš have in common? All of them show the same *relative* variation pattern. All eight subjects have two variants, deletion and retention. And for all eight the deletion variant is the more frequent variant. This is what exemplifies the grammar that the community shares – a relative preference for one variant over another. It is this relative preference that is modeled in the rank-ordering of EVAL. Under this model the claim is that the members of some community will all have the same grammar, and that this grammar will simply stipulate the relative frequency of the variants. In the Latvian case at hand, the community grammar will rank-order the two variants as follows: |deletion ™ retention|. All members of the community

are therefore predicted to prefer deletion over retention, but there can be differences among individuals in how likely they are to select candidates beneath the best candidate as output. This variation between members of the community is, however, not part of their grammar. This is the result of how different individuals interact with the same grammar. By not attempting to model absolute frequencies the rank-ordering of EVAL is not faced with the "which-frequency" problem. In this model it is very clear which frequency to model – the average *relative* frequency that exemplifies the speech community.

The second conceptual problem faced by the three alternative models is the "grammar-alone" problem, and is closely related to the "which-frequency" problem. The alternative models attempt to account through the grammar for the *absolute* frequency with which different variants are observed. An implicit assumption that underlies this is that grammar alone is responsible for the variation pattern. If grammar accounts for all of the variation in the data, then there is no room for other extra-grammatical factors. This goes against a large body of evidence that there are many extra-grammatical factors that significantly interact with variation. It is known, for instance, that factors such as gender, age, socio-economic class, speech situation, individual preferences, *etc*. do contribute towards variation. In the rank-ordering model of EVAL the assumption is that grammar only determines the relative frequency of different variants. The extra-grammatical factors are then responsible for determining the specific absolute frequency with which different variants are observed. Grammar does not account for all aspects of variation, but provides only the limits within which extra-grammatical factors can interplay to determine the absolute patterns of variation.

3.2 Practical problems

In this section I show that crucially unranked constraint grammars and floating constraint grammars are faced with certain practical problems that are avoided by the rank-ordering model of EVAL. The claim of this section is not that the alternatives cannot account for variation data. I will rather show that there are aspects of the variation data that will force the alternatives to construct some unlikely/ad hoc accounts.

The stochastic OT grammars are very powerful and can probably model any variation frequency – since the distance between the basic ranking values of constraints can be specified arbitrarily. However, this is not the case with the crucially unranked constraint grammars and the floating constraint grammars. In order to model a wide array of variation frequencies these kinds of grammars are forced to rely on a large number of constraints. I will illustrate this point here with the example of [t, d]-deletion discussed earlier in this chapter. However, this is a general problem faced by these kinds of grammars and can be illustrated with any variable phenomenon studied in enough different dialects. I use only the deletion rate in pre-consonantal context in the different dialects discussed in §1 above. I list the deletion and retention rates in this context for the different dialects in (41). These data are extracted from the table in (1) in §1.1.

Consider first how these data might be accounted for in a grammar with crucially unranked constraints. In all of the dialects, except for that represented by the Neu data, the deletion variant is the more frequent variant. Excluding the Neu data, it is therefore necessary that the deletion variant be selected as optimal by more of the rankings possible between the crucially unranked constraints than the retention candidate. This means that out of the crucially unranked constraints there should be more constraints that favor

deletion than constraints that favor retention. For the dialects represented by the Neu data, the opposite is true. In these dialects the retention candidate is the more frequent variant. More of the rankings possible between the crucially unranked constraints should therefore favor retention than deletion, implying that out of the crucially unranked constraints there should be more constraints that favor retention than constraints that favor deletion. This implies that it cannot be the same set of constraints that is responsible for the deletion~retention variation in the Neu data and in the other dialects.

(41) **[t, d]-deletion and retention rates in pre-consonantal context in different dialects of English**

	Deletion	**Retention**
Chicano English (Los Angeles)	62	38
Tejano English (San Antonio)	62	38
Black English (Washington, DC)	76	24
Jamaican mesolect (Kingston)	85	15
Trinidadian English	81	19
Neu data	36	64

The picture is more complicated than this, however. Not even for the dialects that all show more deletion than retention can the variation pattern be governed by the same set of crucially unranked constraints. In Chicano and Tejano English the deletion variant is observed 62% of the time. This means that 62% of the rankings possible between the crucially unranked constraints should select the deletion candidate as optimal. However, in Washington, DC AAE deletion is observed 76% of time, so that 76% of the rankings should select the deletion candidate as optimal. In Jamaican English 85% of the rankings

should select the deletion candidate, and in Trinidadian English 81% of the rankings should select deletion. In order to account for all of these different variation patterns, it will be necessary that the set of crucially unranked constraints be different for each of these dialects.

There are two reasons why this situation seems less than ideal. First, it would be very hard to find sets of constraints that would result in exactly the correct variation pattern associated with each dialect. If we take into account that there are many more dialects of English and therefore very likely many more variation patterns, this situation gets just more difficult. It is not in principle impossible to model all of these different variation patterns. However, since so many different sets of constraints would be required, it seems likely that the constraints necessary would be of an ad hoc nature and not very well motivated.

Secondly, we are dealing here with the same phonological process in different dialects of the same language. It is also clear from the data that the grammatical factors that influence the likelihood of application of the process are the same in all of the dialects. If we have to use a different set of constraints to model this process in each dialect, we do not capture the fact that it is the same process applying in each dialect. A significant generalization is then not captured by our analysis.

In a grammar with floating constraints we are faced with much the same problems. There are five different variation patterns represented in the table in (41). One option for modeling this data is to accept that the floating constraint is the same in each of the dialects. Since there are five different patterns, it means that the floating range of the floating constraint will have to be different in each of the five different kinds of dialects.

The floating ranges of different dialects will have to contain different constraints and/or will have to be of different length. Another option is that the floating ranges of different dialects are the same but that the dialects differ in which constraint is the floating constraint. Finally, it is possible that both the floating constraints and the floating ranges are different between dialects. It is probably not in principle impossible to define the constraints that will be necessary to model the different variation patterns. However, in order to model exactly the variation patterns observed it will very likely be necessary to stipulate some rather ad hoc constraints. It will also be the case that the same phenomenon ([t, d]-deletion in pre-consonantal position) is governed by a different set of constraints in different dialects.

These problems are all avoided in the rank-ordering model of EVAL. As shown in the discussion in §1 above, we can use exactly the same two constraints (*Ct#C and MAX) to account for all of the different variation patterns associated with pre-consonantal context. In dialects with more deletion than retention, the ranking ||*Ct#C o MAX|| is observed. EVAL then rank-orders the candidate set as |deletion ™ retention| so that the deletion candidate is the more frequent variant. In dialects represented by the Neu data the constraints are ranked as ||MAX o *Ct#C || so that EVAL imposes the ordering |retention ™ deletion| on the candidate set. Retention is then predicted to be more frequent than deletion.

In the rank-ordering model of EVAL we can model both kinds of dialects with exactly the same constraints – both dialects with more deletion and dialects with more retention. Since the rank-ordering model of EVAL models only relative frequency, all dialects that show more deletion than retention can also be modeled by the same

constraints – it is not necessary to discriminate grammatically between dialects with different absolute preferences for deletion. These differences are the result of how grammar interacts with several extra-grammatical factors. The rank-ordering model of EVAL can therefore account for the variation pattern easily with a small set of well motivated constraints. It can also account for the variation using exactly the same set of constraints in each of the dialects, thereby capturing the fact that it is the same process applying in each of the dialects.

This point has been illustrated here with one specific example. However, it is a general problem that the crucially unranked constraints grammars and floating constraint grammars will be faced with. Any variable phenomenon that applies with different absolute frequencies in different dialects of the same language (or in different languages for that matter), will present these accounts of variation with this same problem. For each dialect a different set of constraints will be necessary to model the variation. The fact that it is the same process conditioned by the same grammatical factors is then not captured. If there happen to be many different dialects with many different absolute variation patterns, the result will be that the same phenomenon has to be governed by many different and distinct sets of constraints. It will then also start becoming increasingly difficult to find exactly the right set of constraints to model each variation pattern.

Appendix: Factorial typologies

A.1 The following phonological context

In 1.2.3.2 I discussed the possible deletion patterns associated with the following phonological context. In this section of the Appendix, I will show that the patterns mentioned there are indeed all and only the patterns predicted under the analysis that I developed there.

A.1.1 No variation patterns

In the rank-ordering model there are two ways in which a categorical phenomenon can be modeled (see Chapter 1 §2.2.3): (i) When all candidates are disfavored by at least one constraint ranked higher than the cut-off, then only the single best candidate is selected as output. (ii) When all but one candidate are disfavored by a constraint ranked higher than the cut-off, then only this one candidate is selected as output.

Consequently, whenever all four constraints outrank the cut-off, no variation will be possible. When the markedness constraint against retention in some context outranks MAX, then that context will show categorical retention. On the other hand, if MAX outranks the markedness constraint against retention for some context, then that context will show categorical retention. In (42) I list all of the possible rankings with the cut-off at the bottom of the hierarchy, as well as the output that is associated with the ranking in each of the three contexts. In this table “D” stands for categorical deletion, and “R” for categorical retention. I indicate cells with more deletion than retention by shading.

(42) **Rankings with the cut-off at the bottom of the hierarchy**

					Pre-C	**Pre-V**	**Pre-P**
*Ct#C	*Ct#V	*Ct##	MAX	Cut-off	D	D	D
*Ct#C	*Ct#V	MAX	*Ct##	Cut-off	D	D	R
*Ct#C	MAX	*Ct#V	*Ct##	Cut-off	D	R	R
MAX	*Ct#C	*Ct#V	*Ct##	Cut-off	R	R	R
*Ct#C	*Ct##	*Ct#V	MAX	Cut-off	D	D	D
*Ct#C	*Ct##	MAX	*Ct#V	Cut-off	D	R	D
*Ct#C	MAX	*Ct##	*Ct#V	Cut-off	D	R	R
MAX	*Ct#C	*Ct##	*Ct#V	Cut-off	R	R	R

Now we can consider the situations in which only one of the candidates for some input is disfavored by a constraint ranked higher than the cut-off. Whenever both MAX and one of the markedness constraints violated by retention in some context ranks below the cut-off, then both retention and deletion for that context will violate no constraints ranked above the cut-off. This context will then show variation. There are therefore two ways in which to assure that only one candidate will violate no constraint ranked above the cut-off: (i) If MAX alone ranks below the cut-off, then only the deletion candidate will violate no constraint above the cut-off, and categorical deletion will be observed in all contexts. (ii) If MAX ranks above the cut-off, then all contexts for which the markedness constraint ranks below the cut-off will show categorical retention – for these contexts only the retention candidate will violate no constraint ranked above the cut-off. (Those contexts for which the markedness constraint also ranks above the cut-off will show either categorical deletion or categorical retention, depending on the ranking between MAX and the markedness constraint applying in the specific context.) In (43) I list all of the rankings that meet one of these two requirements. Comparison of the patterns in (42)

and (43) with those listed in (19a) will confirm that these are indeed all and only the possible patterns with no variation.

(43) **Rankings under which for at least one context only one candidate does not violate a constraint ranked higher than the cut-off**

					Pre-C	**Pre-V**	**Pre-P**
*Ct#C	*Ct#V	*Ct##	Cut-off	MAX	D	D	D
*Ct#C	*Ct##	*Ct#V	Cut-off	MAX	D	D	D
*Ct#C	*Ct#V	MAX	Cut-off	*Ct##	D	D	R
*Ct#C	MAX	*Ct#V	Cut-off	*Ct##	D	R	R
MAX	*Ct#C	*Ct#V	Cut-off	*Ct##	R	R	R
*Ct#C	*Ct##	MAX	Cut-off	*Ct#V	D	R	D
*Ct#C	MAX	*Ct##	Cut-off	*Ct#V	D	R	R
MAX	*Ct#C	*Ct##	Cut-off	*Ct#V	R	R	R
*Ct#C	MAX	Cut-off	*Ct#V	*Ct##	D	R	R
*Ct#C	MAX	Cut-off	*Ct##	*Ct#V	D	R	R
MAX	*Ct#C	Cut-off	*Ct#V	*Ct##	R	R	R
MAX	*Ct#C	Cut-off	*Ct##	*Ct#V	R	R	R
MAX	Cut-off	*Ct#C	*Ct#V	*Ct##	R	R	R
MAX	Cut-off	*Ct#C	*Ct##	*Ct#V	R	R	R

A.1.2 Variation patterns

In (18) I stated the conditions that must be met for variation. Variation in some context can be observed only if both MAX and the markedness constraint that applies in that context rank below the cut-off. Whether the deletion or the retention candidate is the more frequent variant, depends on the ranking between MAX and the markedness constraint. Under the raking ||MAX o Markedness|| the retention candidate is the more frequent variant, and under the ranking ||Markedness o MAX|| the deletion candidate is the more frequent variant. When MAX ranks below the cut-off, then any context whose

markedness constraint ranks above the cut-off will show categorical deletion. In (44) I list all rankings that will result in variation in at least one context. In this table "D" stands for categorical deletion, "R" or categorical retention, "D > R" for variation with more deletion than retention, and "R > D" for variation with more retention than deletion. As before I shade cells with more deletion than retention, i.e. both "D" and "D > R" cells. Comparison of the patterns in (44) with those in (19b) and (19c) will show that these are indeed all and only the patterns with variation in at least one context.

(44) **Rankings under which variation is observed in at least one context**

					Pre-C	**Pre-V**	**Pre-P**
*Ct#C	*Ct#V	Cut-off	*Ct##	MAX	D	D	D > R
*Ct#C	*Ct#V	Cut-off	MAX	*Ct##	D	D	R > D
*Ct#C	*Ct##	Cut-off	*Ct#V	MAX	D	D > R	D
*Ct#C	*Ct##	Cut-off	MAX	*Ct#V	D	R > D	D
*Ct#C	Cut-off	*Ct#V	*Ct##	MAX	D	D > R	D > R
*Ct#C	Cut-off	*Ct#V	MAX	*Ct##	D	D > R	R > D
*Ct#C	Cut-off	*Ct##	*Ct#V	MAX	D	D > R	D > R
*Ct#C	Cut-off	*Ct##	MAX	*Ct#V	D	R > D	D > R
*Ct#C	Cut-off	MAX	*Ct##	*Ct#V	D	R > D	R > D
*Ct#C	Cut-off	MAX	*Ct#V	*Ct##	D	R > D	R > D
Cut-off	*Ct#C	*Ct#V	*Ct##	MAX	D > R	D > R	D > R
Cut-off	*Ct#C	*Ct##	*Ct#V	MAX	D > R	D > R	D > R
Cut-off	*Ct#C	*Ct#V	MAX	*Ct##	D > R	D > R	R > D
Cut-off	*Ct#C	*Ct##	MAX	*Ct#V	D > R	R > D	D > R
Cut-off	*Ct#C	MAX	*Ct#V	*Ct##	D > R	R > D	R > D
Cut-off	*Ct#C	MAX	*Ct##	*Ct#V	D > R	R > D	R > D
Cut-off	MAX	*Ct#C	*Ct##	*Ct#V	R > D	R > D	R > D
Cut-off	MAX	*Ct#C	*Ct#V	*Ct##	R > D	R > D	R > D

A.2 The preceding phonological context

In 2.2.3.2 I discussed the possible deletion patterns associated with the preceding phonological context. In this section of the Appendix, I will show that the patterns mentioned there are indeed all and only the patterns predicted under the analysis that I developed there.

A.2.1 No variation patterns

There are two ways in which a pattern with no variation can be achieved (see Chapter 1 §2.2.3): (i) All candidates violate at least one constraint ranked higher than the cut-off (i.e. all constraints rank higher than the cut-off). (ii) Only one of the candidates violate no constraint ranked higher than the cut-off (i.e. either MAX alone ranks below the cut-off, or MAX ranks above the cut-off and some of the markedness constraints rank below the cut-off).

I begin by considering the first way for achieving a pattern with no variation. The ranking between MAX and the markedness constraint that applies to a specific context will determine whether that context shows categorical deletion or categorical retention. The ranking ||MAX o Markedness|| will result in categorical retention, and the ranking ||Markedness o MAX|| in categorical deletion. In (45) I give one example of such a ranking. Since all constraints rank higher than the cut-off, only the single best candidate for any input will be selected as output. This tableau can therefore be interpreted like a classic OT tableau – i.e. I use the pointing hand in the tableau to indicate the optimal candidate.

(45) **One grammar with the cut-off at bottom of hierarchy**

Context			*[+cor, -son] [+cor, -son]	*[+cor, -cont] [+cor, -cont]	*[-son, -cont] [-son, -cont]	MAX	*[+cor][+cor]	*[-son][-son]	*[-cont][-cont]
[s, š, z, ž]	/st/	-st	*!				*	*	
		-s∅				*			
[n]	/nt/	-nt		*!			*		*
		-n∅				*			
[k, g, p, b]	/kt/	-kt			*!			*	*
		-k∅				*			
[l]	/lt/	-lt					*		
		-l∅				*!			
[f, v]	/ft/	-ft						*	
		-f∅				*!			
[m, ŋ]	/mt/	-mt							*
						*!			

The three contexts for which the markedness constraints rank higher than MAX show categorical deletion. On the other hand, the three contexts for which the markedness constraints rank lower than MAX show categorical retention. The markedness constraints are in a fixed ranking (see (28) above). Since there are six markedness constraints, it follows that there are seven positions for MAX to be ranked into, and therefore that there are seven patterns possible with the cut-off at the bottom of the hierarchy. The seven possible patterns are listed in (46). For each pattern I also mention the highest ranking markedness constraint that is dominated by MAX for that specific

pattern. In this table "D" stands for categorical deletion (all contexts whose markedness constraints outrank MAX), and "R" stands for categorical retention (all contexts whose markedness constraint ranks below MAX). As before, I also shade all cells that represent contexts with more deletion than retention – i.e. all "D" cells.

(46) **All deletion patterns with the cut-off at the bottom of the hierarchy**

MAX ranked above	**[s, š, z, ž]**	**[n]**	**[k, g, p, b]**	**[l]**	**[f, v]**	**[m, ŋ]**
Cut-off	D	D	D	D	D	D
*[-cont][-cont]	D	D	D	D	D	R
*[-son][-son]	D	D	D	D	R	R
*[+cor][+cor]	D	D	D	R	R	R
*[-son, -cont][-son, -cont]	D	D	R	R	R	R
*[+cor, -cont][+cor, -cont]	D	R	R	R	R	R
*[+cor, -son][+cor, -son]	R	R	R	R	R	R

Now we can consider the second source of patterns with no variation. These are rankings in which for some inputs one candidate violates only constraints ranked lower than the cut-off. There are two possibilities. If MAX alone ranks lower than the cut-off, then for all contexts the deletion candidate will violate no constraint ranked higher the cut-off while all other candidates will violate a constraint ranked higher than the cut-off. The deletion candidate will then be the only output observed in all contexts. This is illustrated by the tableau in (47). In this tableau I use the pointing hand like in a classic OT tableau – i.e. to indicate the single output for every input.

(47) **Only MAX below the cut-off = categorical deletion everywhere**

Context			*[+cor, -son] [+cor, -son]	*[+cor, -cont] [+cor, -cont]	*[-son, -cont] [-son, -cont]	*[+cor][+cor]	*[-son][-son]	*[-cont][-cont]	MAX
[s, š, z, ž]	/st/	-st	*!			*	*		
		-s∅							*
[n]	/nt/	-nt		*!		*		*	
		-n∅							*
[k, g, p, b]	/kt/	-kt			*!		*	*	
		-k∅							*
[l]	/lt/	-lt				*!			
		-l∅							*
[f, v]	/ft/	-ft					*!		
		-f∅							*
[m, ŋ]	/mt/	-mt						*!	
									*

There is a second way in which to achieve no variation with some constraints ranked lower than the cut-off, namely if MAX ranks higher than the cut-off and some markedness constraint (or constraints) ranks lower than the cut-off. For all those contexts where MAX ranks higher than the markedness constraint, categorical retention will be observed. This is irrespective of whether the markedness constraint is ranked above or below the cut-off. If the markedness constraint ranks lower than the cut-off, then the retention candidate is the only candidate that does not violate a constraint ranked higher than the cut-off and is therefore selected as only output. If the markedness constraint

ranks higher than the cut-off then all candidates violate some constraint ranked above the cut-off and only the one best candidate is selected as output. If MAX ranks below the markedness constraint for some context, then all candidates again violate some constraint above the cut-off and only the best candidate is selected as output. In these contexts the deletion candidate is therefore selected. The tableau in (48) illustrates one of these grammars. The pointing hand is again used as in classic OT.

(48) **One grammar with some markedness constraint below the cut-off**

Context			*[+cor, -son] [+cor, -son]	*[+cor, -cont] [+cor, -cont]	MAX	*[-son, -cont] [-son, -cont]	*[+cor][+cor]	*[-son][-son]	*[-cont][-cont]
[s, š, z, ž]	/st/	-st	*!				*	*	
		-s∅			*				
[n]	/nt/	-nt		*!			*		*
		-n∅			*				
[k, g, p, b]	/kt/	-kt				*		*	*
		-k∅			*!				
[l]	/lt/	-lt					*		
		-l∅			*!				
[f, v]	/ft/	-ft						*	
		-f∅			*!				
[m, ŋ]	/mt/	-mt							*
					*!				

For the first two contexts we have the ranking ||Markedness o MAX||. Both retention and deletion violate constraints above the cut-off, but deletion violates the lower

ranking constraint. These contexts therefore have categorical deletion. In the third context we have ||MAX o Markedness||. Also here both candidates violate a constraint above the cut-off. Only the best candidate is then selected, which here is the retention candidate. In the last three contexts the markedness constraint ranks below the cut-off and therefore also below MAX. The retention candidate as the only candidate not violating a constraint ranked above the cut-off is selected as only output.

The markedness constraints are in a fixed ranking (see (28) above). Of the constraints that rank higher than the cut-off only MAX can therefore move. If there are *n* markedness constraints higher than the cut-off, then there are (*n* + 1) positions into which MAX can be ranked. In (49) I list all the patterns with no variation that are possible with at least one constraint below the cut-off. The first line represents a grammar in which MAX is the only constraint below the cut-off (see (47)). For each of the other lines I represent the highest ranking markedness constraint that is ranked lower than MAX, as well as the highest ranking markedness constraint ranked lower than the cut-off. In order the determine the output pattern, it is really only necessary to know the highest ranking markedness constraint ranked lower than MAX – if ||MAX o Markedness|| then we have categorical retention and if ||Markedness o MAX|| we have categorical deletion.

Comparison of the patterns in (46) and (49) with the patterns in (33a) will confirm that (33a) does indeed include all and only the no-variation patterns.

(49) **All patterns with no variation and with at least one constraint ranked below the cut-off**

	Constraint that applies	*[+cor, -son] [+cor, -son]	*[+cor, -cont] [+cor, -cont]	*[-son, -cont] [-son, -cont]	*[+cor][+cor]	*[-son][-son]	*[-cont][-cont]
Cut-off above	**MAX ranked above**	**[s, š, z, ž]**	**[n]**	**[k, g, p, b]**	**[l]**	**[f, v]**	**[m, ŋ]**
MAX	–	D	D	D	D	D	D
*[-cont][-cont]	*[-cont][-cont]	D	D	D	D	D	R
	*[-son][-son]	D	D	D	D	R	R
	*[+cor][+cor]	D	D	D	R	R	R
	*[-son,-cont][-son,-cont]	D	D	R	R	R	R
	*[+cor,-cont][+cor,-cont]	D	R	R	R	R	R
	*[+cor,-son][+cor,-son]	R	R	R	R	R	R
*[-son][-son]	*[-son][-son]	D	D	D	D	R	R
	*[+cor][+cor]	D	D	D	R	R	R
	*[-son,-cont][-son,-cont]	D	D	R	R	R	R
	*[+cor,-cont][+cor,-cont]	D	R	R	R	R	R
	*[+cor,-son][+cor,-son]	R	R	R	R	R	R
*[+cor][+cor]	*[+cor][+cor]	D	D	D	R	R	R
	*[-son,-cont][-son,-cont]	D	D	R	R	R	R
	*[+cor,-cont][+cor,-cont]	D	R	R	R	R	R
	*[+cor,-son][+cor,-son]	R	R	R	R	R	R
*[-son, -cont] [-son, -cont]	*[-son,-cont][-son,-cont]	D	D	R	R	R	R
	*[+cor,-cont][+cor,-cont]	D	R	R	R	R	R
	*[+cor,-son][+cor,-son]	R	R	R	R	R	R
*[+cor, -cont] [+cor, -cont]	*[+cor,-cont][+cor,-cont]	D	R	R	R	R	R
	*[+cor,-son][+cor,-son]	R	R	R	R	R	R
*[+cor, -son] [+cor, -son]	*[+cor,-son][+cor,-son]	R	R	R	R	R	R

A.2.2 Variation patterns

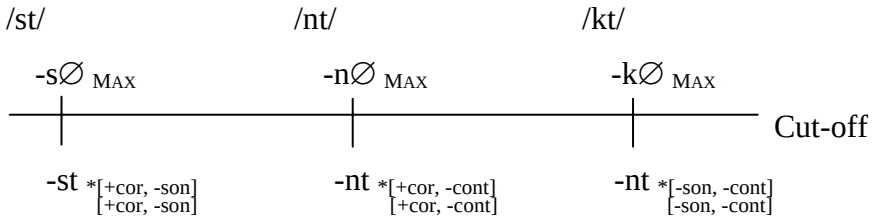
In (32) I listed the conditions that must be met for each context in order for variation to be observed in that context. These conditions can be summarized as follows: For some context to show variation, both the markedness constraints that apply in that context and MAX have to rank lower than the cut-off. The ranking between MAX and the markedness constraints will determine whether deletion or retention will be observed more frequently. Under the ranking ||MAX o Markedness|| more retention than deletion will be observed since the retention candidate violates the lower ranking constraint. Under the ranking ||Markedness o MAX|| more deletion will be observed since deletion violates the lower ranking constraint. Contexts whose markedness constraints rank higher than the cut-off will show categorical deletion – since MAX is below the cut-off the deletion candidate violates no constraint above the cut-off and therefore the retention candidate that does violate a constraint ranked higher than the cut-off will not be accessed as output. The tableau in (50) shows one of these grammars as an example.

For the first three contexts in (50) the ranking ||Markedness o Cut-off o MAX|| holds. The retention candidate violates a markedness constraint ranked higher than the cut-off and the deletion candidate violates MAX ranked lower the cut-off. In these contexts we see categorical deletion. The rank-ordering imposed by EVAL on the candidate set for these three contexts are shown in (51).

(50) **One grammar with variation in at least some contexts**

Context				*[+cor, -son][+cor, -son]	*[+cor, -cont][+cor, -cont]	*[-son, -cont][-son, -cont]	*[+cor][+cor]	MAX	*[-son][-son]	*[-cont][-cont]
[s, š, z, ž]	/st/	2	-st	*!			*		*	
		1	-s∅					*		
[n]	/nt/	2	-nt		*!		*			*
		1	-n∅					*		
[k, g, p, b]	/kt/	2	-kt			*!			*	*
		1	-k∅					*		
[l]	/lt/	2	-lt				*			
		1	-l∅					*		
[f, v]	/ft/	1	-ft						*	
		2	-f∅					*		
[m, ŋ]	/mt/	1	-mt							*
		2						*		

(51) **Output of EVAL for first three contexts from (50)**



The markedness constraints that apply in the last three contexts are all ranked lower than the cut-off. For these contexts neither retention nor deletion violates a constraint ranked higher than the cut-off. We will therefore see variation between deletion and retention in these contexts. For the fourth context, we have the ranking

||Markedness o MAX|| so that the deletion candidate violates the lower ranking constraint. In this context deletion is therefore more frequent than retention. For the last two contexts we have the ranking ||MAX o Markedness|| so that retention violates the lower ranking constraint. For these contexts retention is therefore the more frequent variant. The output of EVAL for these last three contexts is shown in (52).

(52) **Output of EVAL for last three contexts from (50)**

/lt/	/ft/	/mt/
-l∅ $_{\text{MAX}}$	-ft $_{*[-son][-son]}$	-mt $_{*[-cont][-cont]}$
-lt $_{*[+cor][+cor]}$	-f∅ $_{\text{MAX}}$	-m∅ $_{\text{MAX}}$

Cut-off

The ranking between the markedness constraints are fixed (see (28) above). Of the constraints that rank below the cut-off, only MAX can therefore move. If there are *n* markedness constraints ranked below the cut-off, then there are (*n* + 1) positions for MAX to rank into below the cut-off. In (53) I list all the possible patterns with variation. I indicate for each line in the table the highest ranked markedness constraint that is ranked lower than the cut-off. For all those contexts in which higher ranking markedness constraints apply, only deletion will be observed (since the deletion candidate will violate MAX which is ranked lower than cut-off and the retention candidate will violate a markedness constraint ranked higher than the cut-off). Those contexts whose markedness constraints rank lower than the cut-off will show variation (since neither retention nor deletion will violate any constraints above the cut-off). If we have the ranking ||MAX o Markedness|| then the retention candidate will be the more frequent variant (since it violates a lower ranked constraint than deletion). If we have the ranking ||Markedness o

MAX|| then the deletion candidate will be the more frequent variant (since it violates a lower ranked constraint than retention). I therefore also indicate the highest ranked markedness constraint ranked lower than MAX. As before “D” indicates categorical deletion, “R” categorical retention, “D > R” variation with more deletion, and “R > D” variation with more retention. I also shade cells that represent contexts with more deletion than retention, i.e. both “D” and “D > R” cells. Comparison of the patterns in (53) with those in (33b) and (33c) will confirm that (33) does indeed include all and only the possible variation patterns.

(See next page for (53).)

(53) **All patterns with variation**

	Constraint that applies	*[+cor, -son] [+cor, -son]	*[+cor, -cont] [+cor, -cont]	*[-son, -cont] [-son, -cont]	*[+cor][+cor]	*[-son][-son]	*[-cont][-cont]
Cut-off above	**MAX ranked above**	**[s, š, z, ž]**	**[n]**	**[k, g, p, b]**	**[l]**	**[f, v]**	**[m, ŋ]**
*[-cont][-cont]	–	D	D	D	D	D	D>R
	*[-cont][-cont]	D	D	D	D	D	R>D
*[-son][-son]	–	D	D	D	D	D>R	D>R
	*[-cont][-cont]	D	D	D	D	D>R	R>D
	*[-son][-son]	D	D	D	D	R>D	R>D
*[+cor][+cor]	–	D	D	D	D>R	D>R	D>R
	*[-cont][-cont]	D	D	D	D>R	D>R	R>D
	*[-son][-son]	D	D	D	D>R	R>D	R>D
	*[+cor][+cor]	D	D	D	R>D	R>D	R>D
*[-son,-cont] [-son,-cont]	–	D	D	D>R	D>R	D>R	D>R
	*[-cont][-cont]	D	D	D>R	D>R	D>R	R>D
	*[-son][-son]	D	D	D>R	D>R	R>D	R>D
	*[+cor][+cor]	D	D	D>R	R>D	R>D	R>D
	*[-son,-cont][-son,-cont]	D	D	R>D	R>D	R>D	R>D
*[+cor,-cont] [+cor,-cont]	–	D	D>R	D>R	D>R	D>R	D>R
	*[-cont][-cont]	D	D>R	D>R	D>R	D>R	R>D
	*[-son][-son]	D	D>R	D>R	D>R	R>D	R>D
	*[+cor][+cor]	D	D>R	D>R	R>D	R>D	R>D
	*[-son,-cont][-son,-cont]	D	D>R	R>D	R>D	R>D	R>D
	*[+cor,-cont][+cor,-cont]	D	R>D	R>D	R>D	R>D	R>D
*[+cor,-son] [+cor,-son]	–	D>R	D>R	D>R	D>R	D>R	D>R
	*[-cont][-cont]	D>R	D>R	D>R	D>R	D>R	R>D
	*[-son][-son]	D>R	D>R	D>R	D>R	R>D	R>D
	*[+cor][+cor]	D>R	D>R	D>R	R>D	R>D	R>D
	*[-son,-cont][-son,-cont]	D>R	D>R	R>D	R>D	R>D	R>D
	*[+cor,-cont][+cor,-cont]	D>R	R>D	R>D	R>D	R>D	R>D
	*[+cor,-son][+cor,-son]	R>D	R>D	R>D	R>D	R>D	R>D

CHAPTER 6

WELL-FORMEDNESS JUDGMENTS AND LEXICAL DECISION

Language users have the ability to compare different linguistic constructs for their relative well-formedness. In the syntactic literature, for instance, there is a long tradition of marking sentences with different diacritics to indicate their relative well-formedness (*, ?, ??) (Epstein, 1990, 2000, Schütze, 1996). Although relative well-formedness has received less attention in phonology than in syntax, there have been some studies that address this issue for phonology (e.g. Berent and Shimron, 1997, Berent *et al*., 2001a, 2001b, 2002, Frisch *et al*., 2001, Frisch and Zawaydeh, 2001, Hayes, 1997, Hayes and MacEachern, 1996, 1997, Hayes, 1998, Pierrehumbert *et al*., In press).

A very salient property of these judgments is that they always reflect finer distinctions than the categorical distinction between the grammatical (possible) and ungrammatical (impossible). These judgments show a wide array of gradient well-formedness differences between forms. Linguistic theory in the generative tradition has focused strongly on the categorical grammatical/ungrammatical distinction, so that the gradient nature of these judgments has resulted in them receiving relatively little attention in linguistic theory. This is unfortunate, since these judgments are usually strongly influenced by grammar. As such they provide a rich and still largely untapped source of information about the linguistic competence of language users.

The speed with which language users identify non-words in lexical decision tasks is another aspect of linguistic performance that is non-categorical in nature. The speed of

reaction is measured in terms of reaction time, and time is a continuous variable. This measure is therefore necessarily non-categorical.

Lexical decision reaction times have long been used in psycholinguistic studies as a way in which to probe into the workings of the speech processing module (e.g. Balota and Chumbley, 1984, Berent and Shimron, 1997, Berent *et al*., 2001b, Shulman and Davison, 1977, Stone and Van Orden, 1993, Vitevitch and Luce, 1999) . However, again since reaction times are strictly non-categorical while linguistic theories in the generative tradition are categorical, linguistic theory has not paid much attention to lexical decision reaction times as a possible source of information about linguistic competence.

In this chapter of the dissertation I will argue that both well-formedness judgments and reaction times in lexical decision tasks are valuable sources of information about linguistic competence. I will also argue that it is possible to account for these non-categorical phenomena within the rank-ordering model of EVAL.

About well-formedness judgments I will argue that the grammar compares the different non-words in terms of their markedness and imposes a rank-ordering on the non-words. Non-words that receive a higher well-formedness rating are then simply tokens that occupy a relatively high slot in the rank-ordering.

With regard to lexical decision I will argue that language users use the well-formedness of a token as one of the considerations when making lexical decisions. The less well-formed a non-word token is in terms of the grammar of the language, the less seriously the language user will entertain the possibility that the token could be a word, and the quicker the token will be rejected as a non-word. In the rank-ordering model of EVAL, this can be explained as follows: The grammar compares the non-word tokens

and imposes a rank-ordering on them. The lower slot a non-word occupies on this rank-ordering, the less well-formed it is, and the quicker it will be rejected as a non-word.

The rest of this chapter is structured as follows: In §1 I consider some theoretical preliminaries about how the rank-ordering model of EVAL can account for data reported in this chapter. Sections §2 and §3 form the central part of this chapter. In §2 I illustrate how the rank-ordering model of EVAL can account for the relative well-formedness judgments and lexical decision reaction times associated with the Obligatory Contour Principle (OCP) (Goldsmith, 1976, Leben, 1973, McCarthy, 1986, Yip, 1988) in Hebrew. In this section I rely on the data reported by Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997). In §3 I report on a set of experiments in which I replicated Berent *et al.*'s results for a different phonotactic constraint in English. In §4 alternative accounts of the data are considered. Section §5 then contains a summary and conclusion.

1. Preliminary considerations

This section of the chapter is structured as follows: In §1.1 I give an illustration of how the rank-ordering model of EVAL can be used to account for well-formedness judgments and reaction times in lexical decision tasks. In §1.2 I then consider the relationship between lexical statistics and the grammar. There is ample evidence that statistical patterns in the lexicon influence both well-formedness judgments and lexical decision reaction times. Section §1.2 argues that grammar, in addition to these lexical statistics, also contributes towards determining well-formedness judgments and lexical decision reaction times.

1.1 Well-formedness and lexical decision in a rank-ordering model of EVAL

I propose two extensions to the classic OT grammar in this dissertation (see Chapter 1 §1). First, I argue that EVAL can compare not only candidate output forms generated for some input (generated comparison set), but that EVAL can compare any set of forms – even forms that are not related to each other via a shared input (non-generated comparison set). Secondly, I argue that EVAL imposes a harmonic rank-ordering on the full candidate set. EVAL therefore does more than to distinguish the best candidate from the rest. EVAL also imposes a rank-ordering on the non-best candidates. In this section I show how these two extensions to a classic OT grammar enable us to account for well-formedness judgments and lexical decision reaction times. I will use a made-up example for this purpose.

Consider a language L_1 with the ranking ||NoCoda o *Complex o Dep||. In L_1 neither codas nor complex onsets are allowed – both are avoided by epenthesis. Assume that none of the forms [keɪ], [traɪ] or [lud] is an actual word of L_1. Of these three forms only [keɪ] is a possible word of L_1. [traɪ] is not a possible word because it has a complex margin, and [lud] is not a possible word because it has a coda.

Now consider a language L_2 with the ranking ||NoCoda o Dep o *Complex||. Like L_1, L_2 does not allow codas – codas are avoided by epenthesis. However, unlike L_1, L_2 does tolerate complex onsets – because Dep outranks *Complex. Assume that none of the forms [keɪ], [traɪ] or [lud] is an actual word of L_2. Of these three [lud] is not even a possible word, since it has a coda. However, the other two are both possible words.

In (1) these three non-words are compared with each other in each of L_1 and L_2. Next to the tableaux is a graphic representation of the rank-ordering that EVAL will

impose on these three candidates in each language. (On the typographical conventions used in these tableaux, see Chapter 1 §2.2.)

(1) a. **Comparison in L_1**

		NoCoda	*Complex	Dep
1	keɪ			
2	traɪ		*	
3	lud	*		

Output of EVAL

keɪ — Possible word

traɪ *Complex

lud NoCoda

Impossible words

b. **Comparison in L_2**

		NoCoda	Dep	*Complex
1	keɪ			
2	traɪ			*
3	lud	*		

Output of EVAL

keɪ

traɪ *Complex

Possible words

lud NoCoda — Impossible word

In both L_1 and L_2 EVAL imposes the following rank-ordering on these three forms: |keɪ ™ traɪ ™ lud|. If language users had to rate these three forms as to their relative well-formedness, the prediction is that [keɪ] would be rated best, then [traɪ], and finally [lud]. The prediction is therefore that language users will make more than a categorical distinction between possible and impossible words. The L_1-example shows that in addition to the distinction between possible ([keɪ]) and impossible ([traɪ] and [lud]), language users are predicted to also distinguish between the two impossible words in terms of their well-formedness. The L_2-example shows that in addition to the distinction between possible ([keɪ] and [traɪ]) and impossible ([lud]), language users are predicted to also distinguish between the two possible words in terms of their well-formedness.

What about a lexical decision task? The proposal is that language users employ, among other things, the information provided by grammar when making lexical decisions. The more well-formed a token is, the more seriously the language users will consider the possibility that the token might be a word. In general, it is therefore expected that a non-word that is more well-formed will be rejected more slowly than a non-word that is less well-formed. In both L_1 and L_2 the prediction is that [keɪ] will have the slowest rejection time, with [traɪ] faster, and [lud] the fastest. Again, we predict that language users will not only make the categorical distinction between possible and impossible words. In L_1 they will treat different impossible words differently based on their relative well-formedness – although both [traɪ] and [lud] are impossible as words of this language, [lud] is less well-formed and will therefore be rejected faster than [traɪ]. In L_2 they will treat different possible words differently based on their relative well-formedness – although both [keɪ] and [traɪ] are possible as words of this language, [traɪ] is less well-formed and will therefore be rejected more quickly than [keɪ].

In this explanation we use both of the extensions to a classic OT grammar that are argued for in this dissertation. First, about the comparison sets: The three forms compared in (1) are not related to each other via a shared input. If EVAL could only compare candidate output forms generated by GEN for some input, then the comparison in (1) would not even be possible.

Had EVAL only distinguished the best candidate from the rest of the candidates in a comparison set, then the tableaux in (1) would only have given us evidence that [keɪ] is more well-formed than both [traɪ] and [lud]. But we would not have known anything about how [traɪ] and [lud] are related to each other. In L_1 we would then expect language

users to treat the two impossible words the same. And in L_2 we would expect language users to treat the possible word [traɪ] the same as the impossible word [lud]. In both languages the relative well-formedness difference between the two forms [traɪ] and [lud] would not be predicted at all.

In this chapter I will discuss two examples (one from Hebrew and one from English) showing that (i) language users can compare morphologically unrelated forms for their well-formedness, and that (ii) these comparisons are done across more than just two levels. In both examples we find evidence that language users distinguish possible and impossible words in terms of their well-formedness. However, the Hebrew example shows that language users also make well-formedness distinctions within the set of possible words, and the English example shows that language users make well-formedness distinctions in the set of impossible words. This serves as evidence for the two extensions to the classic OT grammar argued for in the dissertation.

1.2 Factors in the processing of non-words

One claim that I will make in this chapter is that language users employ the grammar of their language in the processing of non-words. The response patterns in the gradient well-formedness and lexical decision experiments are therefore predicted to reflect grammatical effects. The claim that grammar influences phonological processing is not novel. For more claims that phonological processing is mediated via grammar see Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997), Brown and Hildum (1956), Coetzee (to appear), Frisch and Zawaydeh (2001), Moreton (2002a, 2002b, 2003), etc. For claims that grammar plays a much smaller role in phonological processing see Coleman and Pierrehumbet (1997), Luce (1986), Newman *et al*. (1997),

Pierrehumbert *et al*. (In press), Treiman *et al*. (2000, Treiman, 1983,), Vitevitch and Luce (1998, 1999), etc.

The factors that have been claimed to influence phonological processing can be classified broadly into two categories, namely lexical and grammatical. In terms of lexical influences, it has been shown that statistical patterns extracted over the lexicon influence phonological processing. In particular, two kinds of lexical statistics that influence phonological processing have been identified, namely *lexical neighborhood density* and *transitional probabilities*. Rather than reviewing the large body of literature about the influence of these two kinds of lexical statistics, I will discuss a few representative examples from the literature.

1.2.1 Lexical statistics

Lexical neighborhood density. The lexical neighborhood density of a token is a function of both the number of words similar to the token and the usage frequency of these words. The more words similar to some form and the more frequent these words are, the denser the lexical neighborhood of this form will be.[1] Both words and non-words have lexical neighborhoods – the lexical neighborhood of a token (word or non-word) is comprised of all those words in the lexicon that are similar to the token to some specified degree. Lexical neighborhood density has been shown to influence the phonological processing of non-words in several ways. Luce and Pisoni (1998), for instance, performed a speeded lexical decision experiment in which they presented their subjects with non-words that differed in lexical neighborhood densities. They found that (i) non-words that occupy

[1] See Luce (1986) and Luce and Pisoni (1998) for an example of exactly how lexical neighborhood density can be calculated. See also §3.2 below.

denser lexical neighborhoods were consistently responded to more slowly than non-words that inhabit sparser lexical neighborhoods, and (ii) non-words that occupy a denser neighborhood were consistently responded to less accurately than non-words from sparser neighborhoods. They interpret these results as follows: The nature of the lexical decision task requires that the percept be checked against entries in the mental lexicon.[2] However, listeners do not check the percept against all entries in the lexicon, but only against entries that are similar to the percept to some specified degree (i.e. against words in the immediate lexical neighborhood of the token). Non-word tokens that inhabit denser lexical neighborhoods therefore need to be checked against more lexical entries than non-words that inhabit sparser lexical neighborhoods, with the result that non-words from sparse lexical neighborhoods can be rejected more quickly than non-words from dense lexical neighborhoods. Similarly, the larger number of lexical entries against which non-words from denser neighborhoods need to be checked results in more opportunities for mistakenly identifying the non-word percept as an actual lexical entry. For more studies that manipulated the lexical neighborhood density of tokens in lexical decision tasks, see *inter alia* Balota and Chumbley (1984), and Vitevitch and Luce (1999).

The results of these lexical decision experiments suggest that denser lexical neighborhoods inhibit the processing of non-words – denser neighborhoods result in slower and less accurate processing. However, several studies have also been reported that seem to present evidence to the contrary – namely that a dense lexical neighborhood

---

[2] This assumption of Luce and Pisoni (that lexical decision by its very nature requires lexical access) is not a necessary assumption. If a percept violates a phonotactic or prosodic constraint that all words in the language must obey, then it can be rejected as a non-word without accessing the lexicon. All of the tokens used by Luce and Pisoni were possible words of English. Their assumption that lexical access is necessary for lexical decision is therefore true for their experimental design. See also Balota and Chumbley (1984).

aid in the processing of non-words. It has, for instance, been shown (i) that non-words from denser lexical neighborhoods are judged more word-like than non-words from sparser neighborhoods (Bailey and Hahn, 1998), and (ii) that non-words from denser neighborhoods are shadowed (repeated) significantly faster than non-words from sparser neighborhoods (Vitevitch and Luce, 1998). It seems to be that a dense lexical neighborhood sometimes aids in the processing of non-words, and sometimes inhibits the processing of non-words. Vitevitch and Luce (1999) argue for a principled distinction between the tasks in which lexical neighborhood density aids processing and the tasks in which lexical neighborhood density inhibits processing. They claim that the inhibitory effect of lexical neighborhood density is observed only in tasks that require lexical access. Lexical decision requires lexical access,[3] and in lexical decision tasks a denser neighborhood counts against a token. However, word-likeness judgments and shadowing do not require lexical access, and in these tasks a denser lexical neighborhood seems to count in favor of a token.

Transitional probability. Based on these results, Vitevitch and Luce (1999) and Newman *et al*. (1997) claim that there are two kinds of lexical statistics that can influence processing of non-words, namely lexical neighborhood density and transitional probabilities. Transitional probability refers to the likelihood of two sounds occurring next to each other in a specific order – i.e. we can compute the probability of, for instance, the sound [k] occurring before the sound [æ] in English. Since tokens with many and frequent neighbors usually also consist of frequent sounds in frequent combinations, there is generally a high positive correlation between the lexical neighborhood density and the

[3] But see the previous footnote.

transitional probabilities of a token – a token with a higher lexical neighborhood density is very likely also to have high transitional probabilities.

Newman *et al.* (1997) and Vitevitch and Luce (1999) claim that information about the transitional probabilities can be accessed without lexical access. If the lexicon is not accessed (as is the case in word-likeness judgments and shadowing tasks), then the lexical neighbors of a token are not activated. A token with many neighbors and a token with few neighbors are therefore equal on this front – neither has to compete with lexical neighbors. In these tasks higher transitional probabilities then aid in the processing of a token. The advantage of tokens with high lexical neighborhoods in these tasks results not from their lexical neighborhoods, but from their transitional probabilities which correlate with their lexical neighborhood density.

Transitional probabilities are calculated in two different ways in the literature. One method is to calculate these probabilities simply over the entries in the lexicon. In this kind of calculation the frequency with which different lexical tokens are used in the language is not taken into consideration in the calculation. To continue with the example used above, in calculating the probability of [k] preceding [æ], a relatively scarce word such as *catamaran* will count equally as much as a relatively frequent word such as *can*. Transitional probabilities are calculated according to this method by, for instance, Frisch and Zawaydeh (2001) for Arabic and Berent *et al*. (2001a) for Hebrew. Another way in which transitional probabilities are calculated, is by weighting the contribution of individual tokens from the lexicon according to their frequency of usage. A more frequent word such as *can* will then contribute more than a less frequent word such as *catamaran*. This method was used for English, *inter alia*, by Vitevitch and Luce (1999),

Pierrehumbert *et al*. (In press), and Treiman *et al*. (2000). Bailey and Hahn (1998:93) calculated transitional probabilities for English according to both methods, and found a very high correlation between the results of the two calculations ($r^2 = 0.96$). Although it seems best to take token frequency into consideration, this result of Bailey and Hahn shows that studies that did not take token frequency into consideration are most likely very comparable to studies that did.[4]

Transitional probabilities have been shown to contribute to phonological processing of non-words in several different ways. Vitevitch and Luce (1998), for instance, performed a speeded repetition task in which they presented their subjects with non-words with high transitional probabilities and non-words with low transitional probabilities. They found that non-words with high transitional probabilities were repeated significantly faster than non-words with low transitional probabilities. Their results therefore show that higher transitional probabilities aid in the processing of non-words.

Transitional probability has also been shown to influence word-likeness or gradient well-formedness judgments. There are many studies that report this result – a few of the more important are Bailey and Hahn (1998), Coleman and Pierrehumbert (1997), Frisch *et al.* (2001), Pierrehumbert *et al*. (In press), Treiman *et al*. (2000). The basic result reported in all of these studies is that a non-word that has higher phoneme

[4] For some languages it is not practically possible to take token frequency into consideration. In order to do to this, information about the usage frequency of different words is required, and this kind of information is for most languages not available. For instance, it is not available for Hebrew or for Arabic. Frisch and Zawaydeh (2001) and Berent *et al*. (2001a) could therefore not take token frequency into consideration.

transitional probabilities is in general judged as more well-formed or more word-like than a non-word with lower transitional probabilities.

There is therefore evidence for the fact that both lexical neighborhood density and transitional probabilities influence the processing of non-words. However, there is also evidence (i) that the influence of these factors are typically rather small (they account for only a small fraction of the variation in the observed data), and (ii) that the influence of lexical statistics is not very robust and can easily be reduced or even completely eliminated by varying the experimental task conditions. In what follows I will briefly present evidence for these two claims before moving on to discussing the evidence for the influence of grammar on the processing of non-words.

Bailey and Hahn (1998) conducted a word-likeness experiment with the express purpose of determining how much transitional probabilities contributed towards explaining the variation in the response data. They collected relative well-formedness judgments on 291 possible non-words of English. They then calculated the weighted phoneme transitional probabilities of the non-words, and performed a regression analysis on the response data with the transitional probabilities of the tokens as the independent variable. They found that transitional probability was significantly correlated with well-formedness judgments. However, they also found that transitional probabilities explained only a very small portion of the variation in their response data ($r^2 = 0.09$, $p < 0.0001$). Although transitional probabilities do contribute towards determining well-formedness judgments, it is clear that they do not contribute much. There must be other factors that also contribute towards the results.

The contribution of transitional probability towards non-word processing is therefore small. In addition to it being small, it is also not a very robust effect. The effect of lexical statistics (both transitional probabilities and lexical neighborhood density) has been shown to decrease or even disappear completely in certain experimental designs. There are at least two situations in which the effects of lexical statistics decrease or disappear. The first is when the experimental setup is such that subjects do not dedicate enough processing resources to the processing of the non-word tokens. Pitt and Samuel (1993), for instance, found that adding a distractor task results in a reduction of the effects caused by lexical statistics. They conducted phoneme identification experiments and found that lexical statistics did influence perception.[5] However, when they added a distractor task (by requiring their listeners to perform rudimentary mathematical calculations during the phoneme identification task), the effect of lexical statistics was significantly reduced.

The second situation in which the effect of lexical statistics can disappear is when the stimulus sets used in experiments are not varied enough. Cutler *et al*. (1987) also conducted phoneme identification experiments. When they used stimulus sets that contain forms of varied prosodic shapes (monosyllabic, disyllabic, initial stress, final stress, etc.), they found that the response patterns were influenced by the lexical statistics.

---

[5] They presented their listeners with tokens ambiguous between two percepts. The one endpoint was a word and the other a non-word. The word endpoint is favored by a lexical bias (Ganong, 1980). Their listeners responded according to this bias – i.e. they were more likely to identify the phoneme such that a real word resulted. However, the advantage of the word endpoint was significantly reduced by the addition of the distractor task.

However, when the stimulus set contained tokens with only one prosodic shape (only monosyllables), the effect of lexical statistics disappeared. [6]

1.2.2 Grammar

Grammar has also been shown to influence the processing of non-words. The general idea behind this claim is that non-words are processed differently depending on how well they conform to grammar. There is evidence that non-words that do conform to the grammar (possible words) are processed differently than non-words that do not conform to the grammar (impossible words). But there is also evidence that grammar makes finer distinctions than simply between possible and impossible words. Even within the set of possible words, a distinction is made between non-words that are more well-formed and non-words that are less well-formed. The same is true for the set of impossible words – also here a distinction between more and less well-formed non-words arises. This has been shown to be true for gradient well-formedness judgments and lexical decision tasks. In the rest of this section I will discuss representative examples from the literature on how grammar influences phonological processing.

Frisch and Zawaydeh (2001) conducted a word-likeness experiment in which they found strong evidence that the grammar (in addition to lexical statistics) contributes towards how word-like language users consider a non-word to be. They investigated the sensitivity of Arabic speakers to the restrictions on the structure of verbal roots. Arabic verbal morphology, like that of other Semitic languages, is based on tri-consonantal verbal roots. The verbal roots are subject to a restriction that prohibits homorganic

---

[6] There are more experimental results reported in the literature that point to the fact that the contribution of lexical statistics to phonological processing is heavily task dependent (Eimas *et al*., 1990, Eimas and Nygaard, 1992, Frauenfelder and Seguí, 1989).

consonants from occurring in the same root (because of some kind of OCP-effect).[7] Frisch and Zawaydeh chose non-words that violate the OCP and non-words that do not violate the OCP. The tokens were chosen such that forms that do violate the OCP had high transitional probabilities and dense lexical neighborhoods, and *vice versa* for forms that obey the OCP. Based on results such as that of Bailey and Hahn (1998) discussed just above, it is therefore expected that the OCP-violating tokens with the higher transitional probabilities and lexical neighborhood densities should be judged as more word-like than the non-violators. However, they found the opposite. The OCP-violating tokens were consistently judged as less word-like in spite of the fact that these tokens were favored by the lexical statistics. This shows (i) that grammar does contribute towards the processing of non-words, and (ii) that the effects of lexical statistics can be overridden by grammar – when there is a conflict between grammar and lexical statistics, grammar wins out.[8]

Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997) conducted a series of experiments in which they report very similar findings for Hebrew.[9] Like Arabic, Hebrew verbal morphology is based on tri-consonantal roots. There are restrictions on the distribution of identical consonants in the tri-consonantal roots. Forms without identical consonants are possible words of Hebrew (i.e. [QiSeM] is a possible word). Also, forms with identical consonants in the positions usually occupied by the final two root

---

[7] The restriction is actually much more complicated than this. See Frisch *et al*. (2004), Greenberg (1950), McCarthy (1986, 1994), Pierrehumbert (1993) for a discussion of the details of the consonant co-occurrence restrictions in Arabic.

[8] See also Berent and Shimron (1997:55-56), Frisch *et al*. (2001:170, 174) and Pierrehumbert (1994:181) about the idea that the effects of grammar can override that of lexical statistics.

[9] The Berent *et al*. studies are discussed in detail in §2.2 below.

consonants are possible words of Hebrew (i.e. [QiSeS] is a possible word). However, forms with identical consonants in the positions usually occupied by the first two root consonants are not possible words of Hebrew (i.e. [QiQeS] is not a possible word).[10] Berent *et al*. conducted a series of experiments to test whether speakers of Modern Hebrew are sensitive to this restriction on possible words. They presented the subjects in their experiments with a stimulus list of non-words, some of which had no identical consonants, some with final identical consonants, and some with initial identical consonants. In the selection of stimuli, they controlled for the possible influence of lexical statistics. The task of the subjects in their experiments was to rate these non-words as to their word-likeness or well-formedness. Berent *et al*. found that possible words (no identical consonants and final identical consonants) were judged better than impossible words (initial identical consonants). However, they also found evidence for a distinction within the set of possible words. Forms with no identical consonants were judged better than forms with final identical consonants – even though both of these forms are possible words of Hebrew. They interpret this result as showing that even forms with final identical consonants violate some well-formedness constraints, and that speakers of Hebrew have access to this information. Grammar therefore makes finer distinctions than simply between possible and impossible words. Grammar also distinguishes within the set of possible words between more and less well-formed tokens.

Berent *et al*. (2001b) have also shown that grammar influences lexical decision. In a lexical decision experiment they presented their subjects with a list of tokens that

---

[10] See Gafos (1998, 2003), McCarty (1986) and Ussishkin (1999) for analyses of this phenomenon. See also §2.1 below for an alternative OT account of this restriction.

contained both words and non-words. They controlled for the possible influence of lexical statistics in the selection of the stimuli. The task of the subjects was to distinguish between words and non-words. Reaction times were measured. Berent *et al*. found evidence for a distinction between possible and impossible words. Impossible non-words (with identical initial consonants) were rejected significantly faster than possible non-words (with identical final consonants). The subjects in their experiment therefore used the information supplied by grammar when making lexical decisions. The less well-formed (actually ill-formed) tokens were considered less seriously as potential words, and therefore rejected more quickly. See also Stone and Van Orden (1993) for results that show that grammar influences lexical decision.

In summary we can say that both lexical statistics and grammar contribute towards the processing of non-words. Therefore, I do not claim that lexical statistics are irrelevant in the processing of non-words, but rather that lexical statistics alone are not sufficient to account for how non-words are processed. We also need to take the contribution of grammar into consideration.

In the next two sections of this chapter I discuss two examples that show that grammar influences the processing of non-words. In §2 I discuss the experiments of Berent *et al*. referred to above in more detail. These experiments show that: (i) Hebrew speakers employ grammar in the processing of non-words. (ii) In particular, these results show that Hebrew speakers make multi-level distinctions in terms of well-formedness. They not only distinguish between possible and impossible words, but that they make distinctions within the set of possible words. Possible words that are more well-formed formed according to the grammar are treated differently from possible words that are less

well-formed. In §3 I then discuss a series of experiments that I conducted in which I replicated the results of Berent *et al*. with English speakers. These experiments show that: (i) English speakers use the information provided by grammar in the processing of non-words. (ii) Specifically, English speakers also make multi-level distinctions. In addition to distinguishing possible and impossible words, they also distinguish between different impossible words in terms of their well-formedness. Impossible words that are more well-formed are reacted to differently than impossible non-words that are less well-formed.

2. The OCP in the processing of non-words in Hebrew

One of the most striking features of Semitic morphology is that the overwhelming majority of verbal roots are tri-consonantal. In addition to this, there are strict limitations on the distribution of consonants between the three consonantal positions in the root (Frisch *et al*., 2004, Gafos, 2003, Greenberg, 1950, McCarthy, 1986, 1994, Morgenbrod and Serifi, 1981, Pierrehumbert, 1993). I will focus here on one of these distributional restrictions, namely on the restriction on the distribution of identical consonants in the root. Forms with identical consonants in the positions usually occupied by the first two root consonants are generally not allowed – i.e. *[QiQeS] is ill-formed. On the other hand, forms with identical consonants in the positions usually occupied by the last two root consonants are well-formed – i.e. [QiSeS] is acceptable. I will refer to forms such as [QiQeS] as initial-geminates, and to forms such as [QiSeS] as final-geminates. (This follows in the tradition of the classical grammars of Hebrew and Arabic in which these forms are called “geminates”.) In this section of the chapter, I will discuss experimental

evidence showing that this restriction has psychological reality for speakers of Modern Hebrew. In particular, it determines how speakers of Modern Hebrew process non-words.

This section of the chapter is structured as follows: In §2.1 I develop an OT analysis of this asymmetrical distribution of identical consonants. The analysis will establish that: (i) forms without identical consonants (neither in initial nor in final consonantal slots), are the most well-formed; (ii) forms with identical consonants in the final two consonantal slots are less well-formed; (iii) and forms with identical consonants in the initial consonantal slots are the least well-formed – i.e. |QiSeM ™ QiSeS ™ QiQeS|. In §2.2 I then discuss the results of a series of experiments performed by Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997) that confirm this harmonic rank-ordering.

2.1 Restrictions on identity in roots

The analysis that I present here follows broadly in the tradition of the OCP-based account suggested by McCarthy (1979, 1981, 1986). The basic assumptions that I make are that (i) geminate forms are derived from bi-consonantal roots, (ii) the final root consonant of bi-consonantal roots spreads in order to create a tri-consonantal stem; (iii) identical contiguous consonants are not allowed in Hebrew roots.

2.1.1 Bi-consonantal roots

McCarthy (1979, 1981, 1986) argued that geminate forms are derived from bi-consonantal roots via autosegmental spreading of the final consonant. Although there has not been agreement about the mechanism involved in copying/spreading of the final root consonant, the idea that these forms are derived from bi-consonantal roots has been

widely accepted in the literature – see for instance Bat-El (1994), Gafos (1998, 2003),[11] and Ussishkin (1999). In the rest of this section I will present evidence in favor of the bi-consonantal root approach. The evidence in favor of the bi-consonantal root comes from two sources – from an active process of denominal verbal formation in Modern Hebrew, and from the results of psycholinguistic experiments with speakers of Modern Hebrew.

Modern Hebrew has an active morphological process of deriving verbs from nominals. Hebrew has many nominals with only two consonants, and verbs are also regularly created from these nominals. This shows that Hebrew grammar should allow for the possibility of creating verbs from bi-consonantal roots. The examples below are from Ussishkin (1999:405). See also Bat-El (1994) for a discussion of this word formation process.

(2) **Verbs derived from bi-consonantal nominals in Modern Hebrew**

Noun/Adjective		**Verb**	
sam	'drug'	*simem*	'to drug, to poison'
dam	'blood'	*dimem*	'to bleed'
mana	'portion'	*minen*	'to apportion'
t^{s}ad	'side'	*t^{s}ided*	'to side'
xad	'sharp'	*xided*	'to sharpen'

There is also evidence from psycholinguistic experiments that speakers of Modern Hebrew can form verbs from bi-consonantal roots. Berent *et al*. (2001a) presented speakers of Modern Hebrew with a fully conjugated verbal exemplar and a bare

[11] Gafos (2003) differs from the other sources cited in that he assumes that the second consonant in the bi-consonantal root is underlyingly specified as geminate (or bi-moraic).

consonantal root (neither the exemplar nor the consonantal root corresponded to actual Hebrew words). The task of the subjects was to conjugate the consonantal root in analogy to the fully conjugated exemplar. Half of the consonantal roots were tri-consonantal and half were bi-consonantal – 48 roots of each kind. Their subjects did not report any particular difficulty in conjugating the bi-consonantal roots, and the responses were in fact quite accurate – more than 93% of the responses were correct. The majority of the correct responses involved repetition of the second consonant.[12]

The denominal verbal derivation and the results of the Berent *et al*. study show that speakers of Modern Hebrew can derive final-geminates from bi-consonantal roots. The grammar of Hebrew should therefore allow for this possibility.

The derivation of geminate verbs from bi-consonantal roots is a necessary part of Hebrew grammar. Therefore, the account offered for the asymmetry in the distribution of identical consonants has to explain why it is always the second consonant that spreads. Abstracting away from the origin of the vowels, the account has to explain why /Q-S/ → [QiSeS] is observed but not /Q-S/ → [QiQeS].

Even though I will argue that geminate verbs are derived from bi-consonantal roots, this does not exclude the possibility of there being roots with identical contiguous

---

[12] Berent *et al*. counted four kinds of responses as correct. (i) Final gemination. (ii) Initial gemination. (iii) No gemination – there are some verbal forms in Hebrew that can be formed with only two consonants – especially when one of the root consonants is "weak" (typically a glide). (iv) Addition – when the subjects added a third consonant that was different from any of the two consonants in the root with which they were presented. When subjects used the addition option, they nearly always added a /y/. A /y/ in a Hebrew root is a "weak" consonant and is often subject to deletion. The responses were distributed as follows between these four categories (as a percentage of the total responses, both correct and incorrect):

Final gemination:	47%	No gemination:	32%
Initial gemination:	< 1%	Addition:	14%

consonants. Richness of the base requires that the grammar be able to derive phonotactically legal output forms from any input (Prince and Smolensky, 1993, Smolensky, 1996). The grammar should therefore also be able to handle input roots such as /Q-Q-S/ and /Q-S-S/. I will argue that, were roots such as these to exist, they will be unfaithfully mapped onto the surface. In particular, one of the identical consonants will be deleted so that the surface correspondent of the root has only two non-identical consonants. The second of the two remaining consonants of the root will then spread, just like in a bi-consonantal root. The result will be that bi-consonantal roots and tri-consonantal roots with identical initial or final consonants all map onto the same surface structure – i.e. /Q-S/, /Q-Q-S/ and /Q-S-S/ all map onto [QiSeS].

When presented with a final-geminate form such as [QiSeS], which of the three possible underlying representations (/Q-S/, /Q-Q-S/ or /Q-S-S/) will a Hebrew speaker assume? I will argue that language users rely on the mechanism of lexicon optimization (Prince and Smolensky, 1993, Smolensky, 1996) in selecting underlying representations. This mechanism will consider different possible underlying representations and select the one that results in the most harmonic mapping from underlying representation to surface form. I will the show that of the three possible underlying representations, the bi-consonantal /Q-S/ results in the most harmonic mapping to the surface form [QiSeS]. Hebrew speakers will therefore assume that the root of any surface geminate is a bi-consonantal form. Although the grammar can handle geminate roots such as /Q-Q-S/ and /Q-S-S/, the Hebrew lexicon does not actually contain such roots.

In the next section (§2.1.2) I will develop an explanation for why bi-consonantal roots are mapped onto tri-consonantal surface forms, putting aside until later (§2.1.3) the

question of why geminate forms are derived from bi-consonantal rather than tri-consonantal roots.

2.1.2 From bi-consonantal roots to tri-consonantal stems

The discussion in the previous section shows that Hebrew grammar has to allow for the possibility that tri-consonantal verbal forms can be derived from bi-consonantal roots. Once this is accepted, we have two questions that need to be answered: (i) Why do the bi-consonantal roots not map onto bi-consonantal output forms? (ii) Why is it never the first consonant of the root that is doubled? I consider each of these questions in turn below.

Verbal stems in Semitic are generally required to end on a consonant (Gafos, 1998, McCarthy and Prince, 1990b).[13] This requirement is expressed by the constraint FINAL-C.[14] In order to see how this constraint can force doubling of the final consonant of a bi-consonantal root, consider the following example: One of the conjugations of

---

[13] This requirement is not that the word should end in a consonant, but the stem. Stem should here be interpreted as that form to which prefixes and suffixes attach. There are two reasons why this constraint cannot refer to the word: (i) Bi-consonantal roots show doubling of the final root consonant even when there are suffixes added to the stem, cf. [maS.Mi.Mim] (Berent *et al*., 2001a:26). The doubling of the final /M/ in this form cannot be ascribed to a need for the word to end in a consonant. (ii) There are many words in Hebrew that end in vowels.

[14] The constraint FREE-V used by Prince and Smolensky in their analysis of Lardil truncation (Prince and Smolensky 1993: Chapter 7, no. (152)) can also be interpreted as a ban on (nominative noun) stems ending in vowels.

Ussishkin (1999:415) argues against FINAL-C on the grounds of it being a "templatic" constraint. He argues that it should be replaced by a constraint that he calls STRONG-ANCHOR-R. This constraint requires the rightmost consonant of the input to have a correspondent at the right edge of the stem. A mapping /Q-S, i-e/ → [Qi.Se] violates this constraint, since the rightmost consonant of the input /S/ does not have a correspondent at the right edge of the stem. This constraint is stated within the schema of the ANCHOR constraints and therefore seems to be a member of a well established family of constraints (McCarthy and Prince, 1995). However, ANCHOR constraints express requirements on the edges of constituents. Ussishkin's constraint crucially cannot refer to the right edge of the input but has to refer to rightmost consonant of the input. His constraint is therefore not really an ANCHOR constraint. In fact, his constraint can be seen as a constraint requiring that the stem ends in a consonant, and stipulating where this consonant comes from. It is therefore a combination of FINAL-C (requiring a consonant at the edge of the stem), and the ranking ||DEP o UNIFORMITY|| (stating that spreading is preferred over epenthesis).

Hebrew is characterized by the vocalic melody *i-e* in the perfect form (this conjugation is known as the *pi'el* in the Hebrew grammatical tradition). The vowels /i-e/ are therefore part of the input of a *pi'el* verb in the perfect. (These vowels can be seen as an inflectional morpheme marking a specific form of the verb.) The input will also contain the consonantal verbal root.[15] The input to form a *pi'el* perfect of the made-up bi-consonantal root /Q-S/ will then be /Q-S, i-e/, and this input has to map onto the output [QiSeS]. In the observed output form the two [S]'s stand in correspondence to a single input /S/. Such a one-to-many correspondence violates the anti-spreading faithfulness constraint INTEGRITY (McCarthy and Prince, 1995). In order to avoid violation of FINAL-C Hebrew opts to violate INTEGRITY. This gives us evidence for the ranking ||FINAL-C o INTEGRITY||. This ranking is motivated in the tableau in (3).

(3) **||FINAL-C o INTEGRITY||**

/Q-S, i-e/	FINAL-C	INTEGRITY
Qi.SeS		*
Qi.Se	*!	

This shows why the faithful candidate is not the optimal candidate. However, there are more candidates than just [QiSeS] that avoid violation of FINAL-C, and we

[15] Both Bat-El (1994) and Ussishkin (1999) argue against the existence of the bare consonantal root. However, there are reasons to believe that bare consonantal roots do exist. (i) Berent *et al.* (2001a) required of the subjects in their experiment to conjugate bare consonantal roots. Their subjects had no difficulty in doing this. (ii) Berent *et al.* (2001a, 2001b, 2002, Berent and Shimron, 1997) and Frisch and Zawaydeh (2001) found evidence for the fact that restrictions that apply specifically to the consonants of the root have psychological reality. This shows that speakers of Hebrew and Arabic do represent the consonantal root as a separate entity.

It is possible that in denominal verbal formation (the focus of Bat-El's and Ussishkin's discussions), the vowels of the nominal root do also form part of the input to the verbal formation process. However, when Hebrew speakers form verbs from roots that are not derived from existing words, and especially when they form verbs from made-up roots, they most likely represent the root as a sequence of bare consonants.

should also rule out these candidates. In the next few paragraphs I will discuss the most important unfaithful competitors of the desired output [QiSeS].

First, we need to account for the fact that FINAL-C is not satisfied by epenthesis of a consonant, i.e. /Q-S, i-e/ does not map onto [QiSeT] with and epenthetic [T]. This candidate will violate the constraint against consonantal epenthesis DEP-C. We also need to consider candidates with different arrangements of the consonants and vowels, i.e. [i.QeS] or [Qi.eS]. These candidates all have onsetless syllables and therefore incur violations of ONSET. Both of these alternative candidates can be ruled out by ranking DEP-C and ONSET higher than INTEGRITY which is violated by the actual output [QiSeS]. This is shown in the tableau in (4).[16]

(4) ||**ONSET o DEP-C o INTEGRITY**||[17]

/Q-S, i-e/	ONSET	DEP-C	INTEGRITY
Qi.SeS			*
QiSeT		*!	
Qi.eS	*!		

Having established the reason for the spreading of one of the root consonants, we now need to establish why it is the final rather than the initial root consonant that spreads – i.e. why does /Q-S, i-e/ not map onto [QiQeS]? I will call on a member of the ALIGNMENT family of constraints (McCarthy and Prince, 1993b) to explain this fact. In

[16] Violation of FINAL-C can also be avoided by deletion of one of the input vowels, i.e. /Q-S, i-e/ → [QiS] or [QeS]. This candidate violates MAX-V or some special version of this constraint against deletion from an affix (Gafos, 1998, 2003, McCarthy and Prince, 1995:370, Ussishkin, 1999:417). These candidates can therefore be eliminated by the constraint MAX-V. However, in the discussion in the text I am focusing on unfaithfulness to the root rather than the affixes, and I will therefore not consider these candidates.

[17] We have no evidence for the ranking between ONSET and DEP-C. However, I follow the principle of ranking conservatism (Chapter 3 §2.3 footnote 9, Chapter 4 §2.1.1), and therefore I rank the markedness constraint ONSET over the faithfulness constraint DEP-C.

particular, on a constraint that requires the left edge of the root to coincide with the left edge of the stem. This constraint is defined in (5).

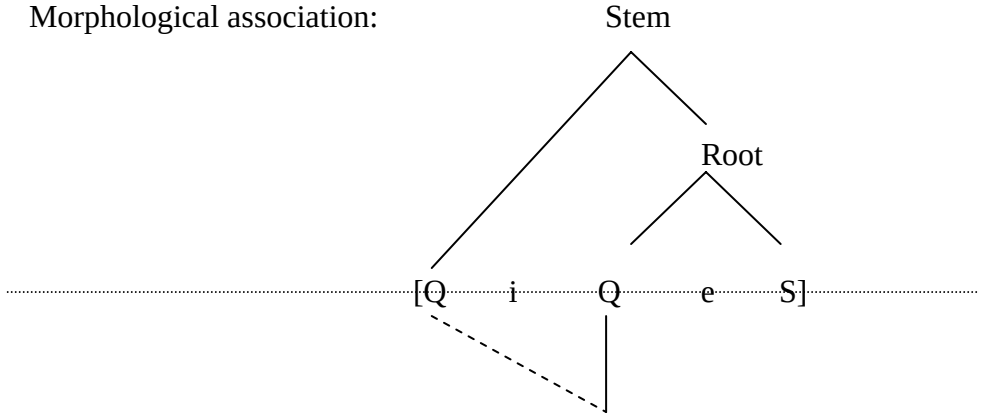
(5) **ALIGN(Root, L, Stem, L)**

The left edge of every root must coincide with the left edge of some stem.

I assume that morphological association and phonological association do not necessarily coincide. Even though the initial [Q] in the output form [QiQeS] is clearly phonologically associated with the /Q/ from the input, it is not morphologically associated with the root morpheme. The initial [Q] in [QiQeS] is a direct daughter of the stem, while the second [Q] is a daughter of the root. The morphological and phonological structure of [QiQeS] is shown in (6).

(6) **Morphological and phonological association in the ungrammatical [QiQeS]**

Morphological association:



Phonological association:

The form [QiQeS] in (6) violates ALIGN(Root, L, Stem, L) (ALIGNL for short) which explains why it is not the observed output. Since both the observed [QiSeS] and the non-observed [QiQeS] violate INTEGRITY, we cannot establish a ranking between ALIGNL and INTEGRITY. However, following the principle of ranking conservatism

(Chapter 4 §2.1.1) I am ranking the markedness constraint ALIGNL higher than the faithfulness constraint INTEGRITY. This is shown in the tableau in (7). In this and all further tableaux I mark morphological association to the root with a superscripted "R" and phonological relatedness with subscripted indexes. Stem boundaries are marked by vertical lines |.

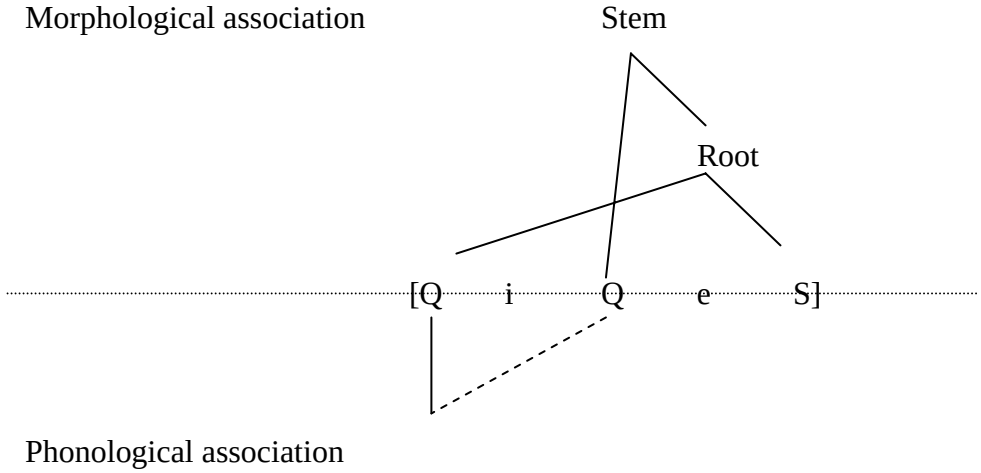
(7) **||ALIGNL ○ INTEGRITY||**

/Q-S, i-e/	ALIGNL	INTEGRITY
$\|Q^R$ i S_i^R e $S_i\|$		*
$\|Q_i$ i Q_i^R e $S^R\|$	*!	*

There are two candidates that obey both FINALC and ALIGNL and that also need to be eliminated. The first root consonant can associate to the left edge of the stem and then to spread onto the second consonantal position of the stem. Similarly, the final root consonant can associate with the right edge of the stem and spread onto the second stem position. These two candidates are represented graphically in (8).

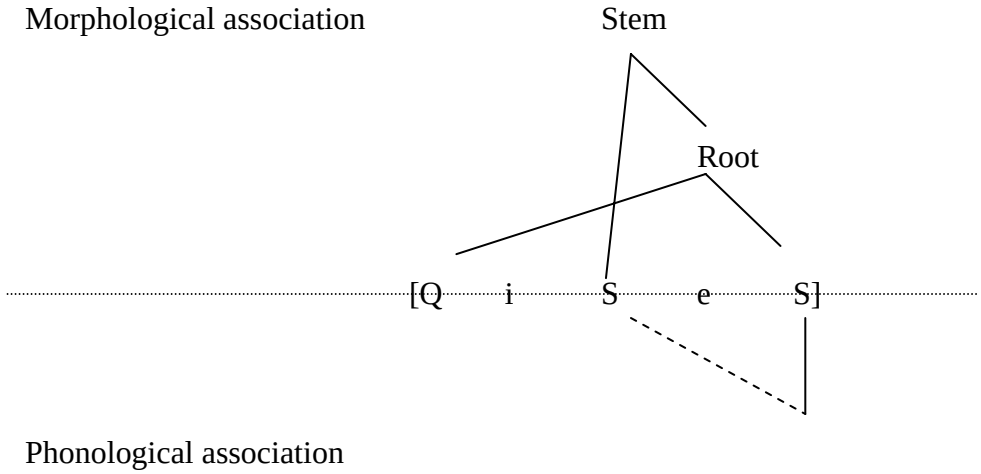
(8) **Two unattested candidates**

a. **Initial root consonant spreads**



((8) continued)

b. **Final root consonant spreads**



Both of these candidates and the observed output candidate [|Q^R i S_i^R e S_i|] violate INTEGRITY. However, both of the forms in (8) also violate the constraint CONTIGUITY (McCarthy and Prince, 1995) indexed to the root – in both of these forms the output stretch that stands in correspondence to the root is discontinuous. This additional violation of $\text{CONTIGUITY}_{\text{Root}}$ is what is responsible for eliminating these candidates.[18] Since both the observed and the non-observed candidates violate INTEGRITY, we cannot establish a ranking between INTEGRITY and $\text{CONTIGUITY}_{\text{Root}}$. This is shown in the tableau in (9).

(9) **The need for $\text{CONTIGUITY}_{\text{Root}}$**

/Q-S, i-e/	$\text{CONTIGUITY}_{\text{Root}}$	INTEGRITY
|Q^R i S_i^R e S_i|		*
|Q^R_i i Q_i e S^R|	*!	*

[18] These forms cannot be ruled out by a constraint against crossing association lines. Roots and stems are presumably on separate tiers, so that these structures can be formed without having the root and stem association lines cross.

In the table in (10) I summarize the rankings that I have argued for in this section. The table includes the ranking and a motivation for the ranking. The number in the last column refers to the crucial examples or the ranking argument that motivates the ranking.

(10) **Summary of the rankings thus far**

Ranking	**Motivation**	**Where?**
FINAL-C o INTEGRITY	Final consonant in bi-consonantal roots doubles	(3)
DEP-C o INTEGRITY	FINAL-C satisfied by spreading, not epenthesis	(4)
ONSET o INTEGRITY	FINAL-C satisfied by spreading, not re-arranging vowels and consonants	(4)
FINAL-C, ONSET, ALIGNL o DEP-C, INTEG, $\text{CONTIG}_{\text{ROOT}}$	Ranking conservatism	

2.1.3 The OCP: banning identical consonants

In this section I will first show that there is a constraint that bans identical contiguous consonants from the surface realization of a root (§2.1.3.1). After that I will argue that there is also a constraint against contiguous identical consonants in the stem, but that Hebrew tolerates violation of this constraint (§2.1.3.2).

2.1.3.1 No identical consonants in the surface realization of the root

The tendency to ban co-occurrence of identical structures within a specific domain is well established cross-linguistically. Leben (1973) and Goldsmith (1976) originally formulated the Obligatory Contour Principle (OCP) to account for the avoidance of

identical contiguous tones. However, the application of the OCP has since been extended also to segmental phonology (McCarthy, 1979, 1986, Padgett, 1991, Yip, 1988).

I will argue here that a specific instantiation of the OCP is responsible for banning contiguous identical consonants from roots in Hebrew. This constraint is defined in (11).

(11) **$\mathbf{OCP_{Root}}$**

Do not allow identical contiguous consonants in the surface realization of a root.

Note that this constraint is formulated such that it is evaluated on the surface – it does not evaluate the underlying representation of a morpheme, but rather the surface realization of a morpheme. It is like all other markedness constraints in OT a surface oriented constraint. This is an important distinction. Had the constraint been formulated to ban identical contiguous consonants from the underlying representation of a morpheme, it would have been in effect a morpheme structure constraint (Chomsky and Halle, 1968), placing a limitation on what can count as a possible input to the grammar. This would not have been in agreement with the principle of "richness of the base" (Prince and Smolensky, 1993, Smolensky, 1996), one of the central tenets of OT. Richness of the base requires that the grammar be able to handle any input. As I will illustrate below (§2.1.6), OCP_{Root} will via lexicon optimization (Prince and Smolensky, 1993) result in a lexicon that contains no roots with identical contiguous consonants. But the grammar will still be able to map a hypothetical morpheme with identical contiguous consonants onto a licit output.

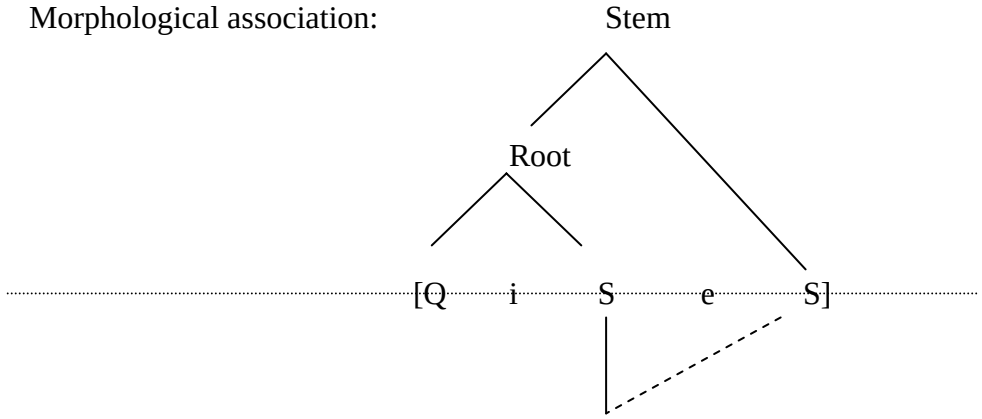
It is important that the domain over which the contiguity of two consonants is determined should be the root. Root consonants are often separated from each other by

intervening non-root vowels. If contiguity were determined directly over the surface form, then two root consonants separated by non-root vowels would not be contiguous. In a form such as [$Q^R iS^R eS^R$], the two [S]'s are contiguous in the root even if they are not directly contiguous on the surface. This form thus violates OCP_{Root} even though the two [S]'s are not strictly contiguous.

What is the evidence that the domain over which the Hebrew constraint is evaluated is the root rather than some other morphological or prosodic domain? One part of the evidence rests on the assumption that the phonological and morphological structures of geminate verbs are not isomorphic. Identical consonants do occur in a geminate verb, but the second of the identical consonants is morphologically associated with the stem and not with the root. A form such as [QiSeS] then has the morpho-phonological structure shown in (12). This form does not contain identical consonants within the surface realization of the root. It does contain identical consonants within the surface realization of the stem. Identical consonants are therefore tolerated in the stem.

(12) **Morphological and phonological association of the final-geminate [QiSeS]**

Morphological association:



Phonological association:

In addition to allowing identical contiguous consonants within the stem, Hebrew also allows identical contiguous consonants within the word. I assume that the stem is the structure to which prefixes and suffixes attach. In (13) I give some examples of identical consonants that are separated by a stem boundary (stem boundaries are indicated by a vertical line |).

(13) **Identical consonants tolerated across a stem boundary in Hebrew**

/ti	+	tfor/	→	[ti|tfor|]	‘she will sew’	
3 f.s.		sew				(Berent *et al*., 2001a:24)
/li	+	lboš/	→	[li|lboš|]	‘to wear’	
inf. marker		wear				(Berent *et al*., 2001a:24)
/rokem	+	im/	→	[|rokm|im]	‘they sew’	
sew		m. pl.				(Uri Strauss, p.c.)

Contiguous identical consonants are disallowed in the root, but tolerated in the stem and in the word. This serves as motivation for the fact that the domain of the relevant OCP-constraint is indeed the root.

Where does the constraint OCP_{Root} fit into the constraint hierarchy for Hebrew? Suppose that there were a root with two initial identical consonants in its underlying representation, i.e. /Q-Q-S/. Since Hebrew does not tolerate violation of OCP_{Root}, this root cannot be mapped onto a surface realization in which both of the underlying /Q/’s are faithfully preserved. One of the underlying /Q/’s has to be treated unfaithfully. There are several ways in which this can be done. One way is by deleting one of the offending

/Q/'s and then treating the root like a bi-consonantal root – i.e. spreading the final /S/. If this repair is used, then both MAX and INTEGRITY are violated. Both of these faithfulness constraints therefore have to rank lower than OCP_{Root}, i.e. ||OCPRoot o {MAX-C, INTEGRITY}||.[19] This is illustrated in the tableau in (14). I do not consider the vowels in the candidates in this tableau.

(14) **||OCP_{Root} o {MAX-C, INTEGRITY}||**

/Q-Q-S/	OCP_{Root}	MAX-C	INTEGRITY
$Q^R Q^R S^R$	*!		
$Q^R S_i^R S_i$		*	*

The tableau in (14) shows that OCP_{Root} is not a morpheme structure condition in the traditional sense of the word (Chomsky and Halle, 1968). It does not limit what counts as a possible input to the grammar. A root with identical contiguous consonants in underlying form is allowed. However, such a morpheme will never be mapped faithfully onto the surface.

2.1.3.2 Avoiding identical consonants in the stem

If the OCP can be indexed to one morphological category (the root), then it should in principle be possible to index it to other morphological categories such as the stem and the word. In addition to the constraint OCP_{Root} defined above, there will then also be

---

[19] This is, of course, not the only way in which violation of OCP_{Root} can be avoided. It is also possible that one of the offending /Q/'s is simply parsed unfaithfully by changing its place of articulation, i.e. /Q/ → [T]. This will then result in a mapping /Q-Q-S/ → [T-Q-S] that also does not violate OCP_{Root}. This would require the ranking ||OCP_{Root} o IDENT(place)||.

Since the Hebrew lexicon does not contain a root like /Q-Q-S/, we cannot know for sure which of the possible repairs Hebrew would use. In the discussion here I am assuming the deletion-plus-spreading repair, simply because I already need spreading as a way in which to satisfy FINAL-C. However, this is to some extent an arbitrary choice. The same point could be made by assuming the repair used to avoid a violation of OCP_{Stem} is violation of an IDENT constraint.

other OCP-constraints such as OCP_{Stem} and OCP_{Word}.[20] I will use OCP_{Stem} as an example in the discussion below.

(15) **OCP_{Stem}**

Do not allow identical contiguous consonants in the surface realization of a stem.

The ranking of OCP_{Root} in Hebrew has been determined above. But what about OCP_{Stem}? Unlike OCP_{Root}, Hebrew does tolerate violation of OCP_{Stem}. This rests on the assumption that the morphological and phonological structure of final-geminate verbs are as represented in (12) above. The identical consonants in final-geminates are both affiliated with the stem, and these final-geminates therefore violate OCP_{Stem}.

Since Hebrew tolerates violation of OCP_{Stem} but not of OCP_{Root}, we can infer the ranking $||OCP_{Root} \circ OCP_{Stem}||$. From the fact that Hebrew tolerates violation of OCP_{Stem}, we can also infer that all faithfulness constraints that could be violated to avoid a violation of OCP_{Stem} should rank higher than OCP_{Stem}. Of the faithfulness constraints considered above only DEP-C can be violated in order to avoid an OCP_{Stem} violation. Suppose that FINAL-C is satisfied not by spreading but by epenthesis of a consonant. The input /Q-S, i-e/ will then map onto a candidate [QiSeT] where the [T] is not phonologically related to any of the consonants in the root. The structure of this candidate is shown in (16).

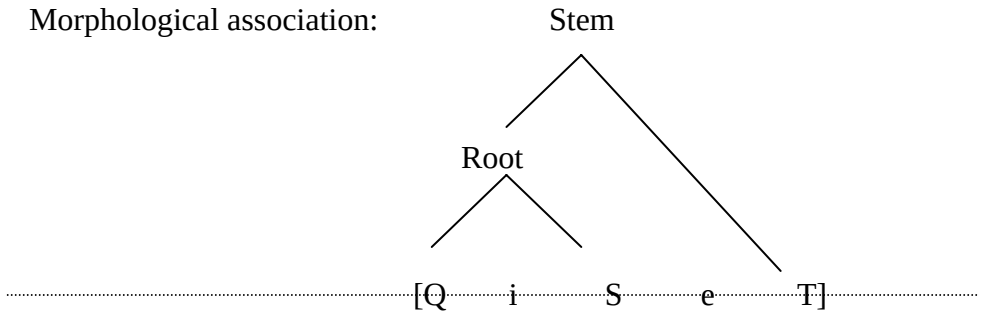
Since Hebrew tolerates violation of OCP_{Stem}, this implies that DEP-C should outrank OCP_{Stem}, i.e. $||$DEP-C $\circ$ $OCP_{Stem}||$.[21] This is shown in the tableau in (17).

---

[20] There is considerable cross-linguistic evidence for consonantal co-occurrence constraints that apply to different morphological domains. See Tessier (2003, 2004) for a recent review of the data.

(16) **Epenthesis to avoid violation of OCP_{Stem}**

Morphological association:



Phonological association:

(17) **||DEP-C o OCP_{Stem}||**

/Q-S/	DEP-C	OCP_{Stem}
$\|Q^R S^R S\|$		*
$\|Q^R S^R T\|$	*!	

The existence of OCP_{Stem} means that final-geminate verbs, even though they are tolerated in Hebrew, are marked relative to verbs derived from roots with three different consonants. A mapping such as /Q-S-M, i-e/ → [|QiSeM|] does not violate OCP_{Stem}, while a mapping such as /Q-S, i-e/ → [|QiSeS|] does. We should see the effects of OCP_{Stem} *inter alia* in the processing of non-words. Even tough both [|QiSeM|] and [|QiSeS|] are possible words of Hebrew, [|QiSeS|] is more marked, and should therefore be rated as less well-formed and should be rejected more quickly as a non-word in a lexical decision task.

---

[21] Strictly speaking there is also a situation in which MAX-C in conjunction with DEP-C could be violated in order to avoid a violation of OCP_{Stem}. Consider a root /Q-S-S/ that is mapped onto $[Q^R S^R T]$. In this output, one of the offending /S/'s has deleted in violation of MAX-C. A [T] has then been inserted in violation of DEP-C. However, this is already ruled out but the ranking ||DEP-C o OCP_{Stem}||.

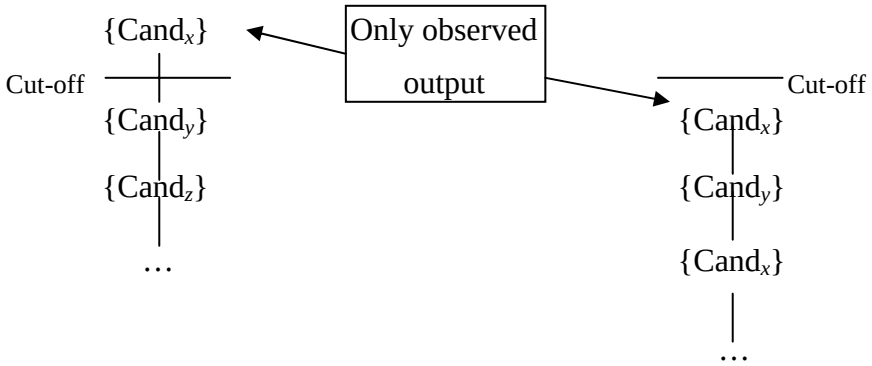
2.1.4 The critical cut-off

The basic theoretical claim that I make in this dissertation is that EVAL does more than just to distinguish between the winning candidate and the mass of losers. EVAL imposes a harmonic rank-ordering on the full candidate set. Language users can then access the full candidate set via this rank-ordering. This idea was used to explain variation in phonological production (Chapters 3, 4 and 5). The most frequently observed variant corresponds to the candidate occupying the highest slot in the rank-ordering, the second most frequent variant to the candidate in the second slot of the rank-ordering, etc. However, if language users have access to the full rank-ordered candidate set, why is variation limited? For most inputs no variation is observed. And even in those instances where variation is observed, it is usually strictly limited to two or three variants. In order to account for the strict limits on variation, the concept of the critical cut-off was introduced (Chapter 1 §2.2.3). The critical cut-off is a point on the constraint hierarchy. If given a choice, language users will not access candidates eliminated by constraints ranked higher than the critical cut-off. Variation is limited because most candidates are eliminated by constraints ranked above the cut-off.

But what of a system with no variation? There are two ways in which an input can be invariantly mapped onto a single output. Suppose that for some input all output candidates except for one is eliminated by a constraint ranked higher than the cut-off. The single candidate that is not eliminated before the cut-off is reached will then be the only observed output candidate – this is just variation with a single observed variant. But there is another way in which an input can be mapped onto a single output. If all candidates violate at least one constraint ranked higher than the cut-off, then language users will be

forced to access candidates eliminated by constraints ranked above the cut-off. However, in order to minimize access of candidates eliminated above this cut-off, language users will access only one candidate, namely the candidate occupying the top slot in the rank-ordering. The two ways in which a categorical phenomenon can be modeled are represented graphically in (18). In this representation the rank-ordered candidate set is represented with the candidate rated best at the top. A horizontal line indicates the location of the critical cut-off. Candidates below this line are eliminated by constraints ranked higher than the cut-off.

(18) **Categorical phenomena in a rank-ordering model of EVAL**



This example shows that one way in which to achieve a situation with no variation, is to rank all constraints above the critical-cut-off. All candidates are then guaranteed to violate at least some constraint ranked higher than the cut-off. (This scenario is represented on the right in (18).) Since variation in production is the exception rather the rule, the assumption should be that our grammar should be constructed so as to allow for variation only when there is explicit evidence of variation. For this reason, I made the conservative assumption that the critical cut-off is located as low as possible on the constraint hierarchy when I discussed variation in production (Chapter 4 §3.1).

We have no evidence of variation in the production of Hebrew geminate verbs. Based on the discussion in the previous paragraph it therefore follows that the critical cut-off should be located right at the bottom of the hierarchy for Hebrew. For any input, all output candidates will then violate some constraint ranked higher than the cut-off, so that only one candidate will be accessed for any input.

2.1.5 Summary

In the table in (19) I list all the rankings that I have argued for thus far. Part of this table is a repetition of the table in (9). However, it also contains information in the rankings that were introduced since (9). The table contains the ranking, a short motivation for the ranking, and where in the preceding discussion that ranking was discussed. After the table I give a graphic representation of the rankings.

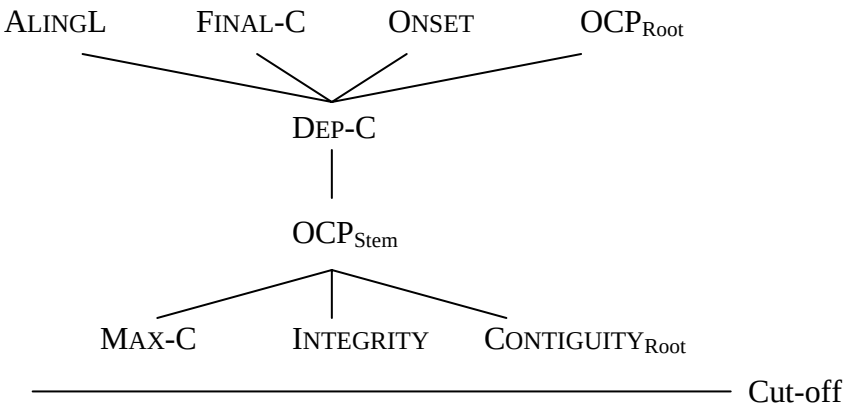
(19) **Summary thus far**

Ranking	**Motivation**	**Where?**
FINAL-C o INTEGRITY	Final consonant in bi-consonantal roots doubles	(3)
DEP-C o INTEGRITY	FINAL-C satisfied by spreading, not epenthesis	(4)
ONSET o INTEGRITY	FINAL-C satisfied by spreading, not re-arranging vowels and consonants	(4)
FINAL-C, ONSET, ALIGNL o DEP-C, INTEGRITY	Ranking conservatism	
OCP_{Root} o MAX-C, INTEGRITY	Geminate roots unfaithfully mapped	(14)
OCP_{Root} o OCP_{Stem}	OCP_{Stem} violated but OCP_{Root} not	§2.1.3.2

((19) continued)

Ranking	**Motivation**	**Where?**
DEP-C o OCP_{Stem}	OCP_{Stem} violation not avoided by epenthesis	(17)
OCP_{Root} o DEP-C	Ranking conservatism	
OCP_{Stem} o INTEG, MAX-C, $CONTIG_{Root}$	Ranking conservatism	
All constraints o Cut-off	No variation, ranking conservatism	§2.1.4

(20) **Graphic representation of the rankings for Hebrew**



In the rest of this section I will illustrate how the hierarchy in (20) generates only actually observed forms for Hebrew verbs. In the next section (§2.1.6), I will then show how this hierarchy together with lexicon optimization ensures that the Hebrew lexicon will not contain roots with identical contiguous consonants.

Let us begin by considering the simplest case – a verbal root with three non-identical consonants. In such a form all three root consonants should be faithfully parsed onto the surface, i.e. /Q-S-M, i-e/ → [QiSeM]. The tableau in (21) considers this case. In

this and all further tableaux I do not consider the vowels. The tableaux from here on should also be interpreted like tableaux in the rank-ordering model of EVAL – see Chapter 1 §2.2.1 and §2.2.3 for the conventions used in these tableaux.

(21) **/Q-S-M/ → [Q S M]**

/Q-S-M/	ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$CONTIG_{Root}$
1 $\lvert Q^R\ S^R\ M^R\rvert$									
3 $\lvert Q^R\ S^R\ T\rvert$	*!				*			*	
2 $\lvert Q^R S^R M^R{}_i M_i\rvert$	*!					*	*		

Output of EVAL[22]

$\lvert Q^R\ S^R\ M^R\rvert$ — Cut-off

Other candidates — Some faithfulness constraint

Since the faithful candidate does not violate any of the markedness constraints, any faithfulness violation is guaranteed to be fatal. All candidates appear below the line representing the critical cut-off. Only one candidate is therefore accessed, namely the faithful candidate.

Now consider how a bi-consonantal root will be handled in this grammar. Such an input has to be mapped onto a form with a final-geminate, i.e. /Q-S, i-e/ → [QiSeS]. The tableau in (22) shows that our grammar does indeed select the candidate as output for such as input.

[22] Candidate [$\lvert Q^R\ S^R\ M^R\rvert$] violates none of the constraints in this mini-grammar. However, it does violate several constraints that are just not considered here. This is why this candidate still occurs below the critical cut-off on this representation.

(22) **/Q-S / → [Q S S]**

/Q-S/	ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$CONTIG_{Root}$
1 $\lvert Q^R\ S_i^R S_i \rvert$						*	*		
5 $\lvert Q_i\ Q_i^R\ S^R \rvert$	*!					*	*		
4 $\lvert Q^R\ S^R \rvert$		(*)[23]	(*)[23]						
3 $\lvert Q^R\ S^R\ T \rvert$					*!				
2 $\lvert Q^R\ S_i\ S_i^R \rvert$						*	*		*!

Output of EVAL

$\lvert Q^R\ S_i^R S_i \rvert$ — Cut-off

$\lvert Q^R\ S_i\ S_i^R \rvert$ CONTIGUITYRoot

Other candidates DEP-C or higher

In the second candidate [$\lvert Q_i\ Q_i^R\ S^R \rvert$] the root and stem are misaligned at their left edges. This leads to a fatal violation of ALIGNL. In the third candidate [$\lvert Q^R\ S^R \rvert$] there are not enough consonants to satisfy both FINAL-C and ONSET simultaneously. This candidate actually stands in for two candidates here, namely [$\lvert Q^R$ i . S^R e|] and [Q^R i . e S^R|]. The first violates FINAL-C and the second ONSET. The fourth candidate [$\lvert Q^R\ S^R\ T \rvert$] inserts a consonant, thereby earning a violation of DEP-C. The last candidate $\lvert Q^R\ S_i\ S_i^R \rvert$ deletes on the two root /S/'s, and then spreads the remaining /S/ into the middle stem position (this is the candidate from (8b) above). The portion of the surface string that stands in correspondence to the root is not contiguous, which earns this candidate a fatal violation of $CONTIGUITY_{Root}$. The first candidate [$\lvert Q^R\ S_i^R S_i \rvert$] is then selected as

[23] See the discussion just below the tableau about these violations.

output in spite of its OCP_{Stem} and INTEGRITY violations. All candidates violate at least one constraint ranked higher than the cut-off. Only the single best candidate is therefore predicted to be accessed – i.e. no variation is predicted.

What remains to show now is how this grammar will deal with a root that has two contiguous identical consonants. I will show just below that lexicon optimization will prevent such roots from being added to the lexicon of Hebrew. However, richness of the base requires that the grammar at least be able to handle such inputs. The tableau in (23) shows how the grammar will handle a root with identical second and third consonants.

(23) **/Q-S-S/ → [$Q^R S_i^R S_i$]**

/Q-S-S/	ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$CONTIG_{Root}$
1 $\|Q^R S_i^R S_i\|$						*	*	*	
3 $\|Q^R S^R S^R\|$				*!		*			
4 $\|Q_i Q^R S^R\|$	*!					*	*	*	
2 $\|Q^R S^R T\|$					*!			*	

Output of EVAL

Cut-off

$|Q^R S_i^R S_i|$ OCPStem

$|Q^R S^R T|$ DEP-C

Other candidates ALIGNL or OCPRoot

The second candidate [$|Q^R S^R S^R|$] is the faithful candidate. It fatally violates OCP_{Root}. The third candidate [$|Q_i Q^R S^R|$] avoids violation of OCP_{Root} by deleting one of the offending /S/'s and then spreading the /Q/ into the initial position of the stem. Since this initial [Q] is not morphologically associated with the root, this candidate does not

violate OCP_{Root}. However, at the same time the disassociation between the initial [Q] and the root means that the left edges of the root and the stem do not coincide. This candidate therefore fatally violates ALIGNL. The last candidate [$|Q^R\ S^R\ T|$] also avoids violation of OCP_{Root} by deletion one of the two [S]'s. In order to satisfy FINAL-C it then inserts a third consonant that is unrelated to any of the input consonants. This earns it a fatal violation of DEP-C. This leaves only the first candidate [$|Q^R\ S_i^R\ S_i|$]. This candidate also avoids violation of OCP_{Root} by deleting one of the offending /S/'s. However, this candidate spreads the remaining /S/ into the final stem position, earning it a violation of INTEGRITY and OCP_{Stem}. Both of these constraints rank lower the constraints violated by other candidates. The first candidate is therefore rated best by EVAL. Again, since all candidates violate at least one constraint ranked higher than the cut-off, no variation is predicted. The result is that a root with identical contiguous consonants will never be mapped faithfully. The situation is the same with a root with identical first and second consonants.

Our grammar does therefore correctly account for the surface patterns observed in Hebrew. Recall that the goal at the outset of this section was to explain the asymmetry in the distribution of identical consonants in Hebrew verbs. In particular, that final-geminates [QiSeS] are well-formed while initial-geminates [QiQeS] are not. How does the grammar developed here account for this fact? There are two possible sources of geminates – they can either be derived from bi-consonantal roots or from tri-consonantal roots with identical consonants. Consider first the bi-consonantal roots. We have to explain why /Q-S/ maps onto [$|Q^R\ S_i^R\ S_i|$] and not onto [$|Q_i\ Q_i^R\ S^R|$]. Both of these forms violate INTEGRITY and OCP_{Stem}. However, the stem and root of [$|Q_i\ Q_i^R\ S^R|$] are

misalignment at their left edges, earning this form an additional violation in terms of ALIGNL.

Now consider the second possible source of geminates – tri consonantal roots with contiguous identical consonants (/Q-Q-S/ and /Q-S-S/). Because of the high ranking of OCP_{Root} the surface realization of a root is not allowed to contain two contiguous identical consonants. In order avoid violation of OCP_{Root}, one of the identical consonants is deleted – hence the ranking ||OCP_{Root} o MAX-C||. The remaining two root consonants are then treated just like a bi-consonantal root – the final root consonants spreads.

2.1.6 Lexicon optimization

When a Hebrew speaker is presented with a word that has two contiguous identical consonants, what prevents him/her from inferring that this word is derived from a root with identical consonants? More concretely, if a Hebrew speaker were to hear a word like [QiSeS], why will he/she assume that this word is the result of an unfaithful mapping from /Q-S, i-e/ rather than the result of the faithful mapping from /Q-S-S, i-e/? The principle of "lexicon optimization" (Prince and Smolensky, 1993) is responsible for this.

Lexicon optimization is proposal for how language users acquire the entries of their lexicon. The idea is that upon perceiving a word [output], language users consider all possible /input/-forms that could be mapped onto [output] by their grammar. They then select from among all the possible /input/-forms that one that results in the most harmonic /input/ → [output] mapping. In lexicon optimization the output candidate is kept constant and the input is varied. Lexicon optimization is then a comparison between different inputs for the same output candidate rather than a comparison between different output candidates for the same input.

The Hebrew example that we are dealing with here is slightly more complicated than this. The output form [QiSeS] can have more than one morphological and phonological structure. For illustration I list just a few of the more obvious structures: [|Q^R i S^R e S^R|] (perfect alignment or root and stem, completely faithful), [|Q^R i S_i^R e S_i|] (root and stem misaligned at right edge, second [S] is the result of spreading from the first), [|Q^R i S^R e S|] (root and stem misaligned at the right edge, second [S] is result of epenthesis), etc. All of these different morpho-phonological forms have the same phonetic interpretation (are pronounced exactly the same). Upon hearing [QiSeS] a Hebrew speaker therefore has to do more than simply consider different possible inputs for this percept. He/she also has to consider different possible morphological and phonological structures for the percept.

I propose that Hebrew speakers will proceed as follows: First, they will feed /Q-S-S, i-e/ and /Q-S, i-e/ separately through their grammars – i.e. a straightforward production oriented process. From each of these two comparisons they will then select the best /input/ → [output] mapping. Lexicon optimization is achieved by comparing the best /input/ → [output] mappings from each of these two comparisons. In (22) and (23) above I showed how the Hebrew grammar that I have developed will treat a /Q-S, i-e/ and /Q-S-S, i-e/ input. The most harmonic mapping from tableau (22) is /Q-S, i-e/ → [|Q^R i S_i^R e S_i|] and the most harmonic mapping from tableau (23) is /Q-S-S, i-e/ → [|Q^R i S_i^R e S_i|]. Lexicon optimization is performed by comparing these two mappings with each other. This comparison is shown in (24).

(24) **Lexicon optimization: [QiSeS] from /Q-S/ not /Q-S-S/**

		ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$CONTIG_{Root}$
2	/Q-S-S/ → [|Q^R $S_i^R S_i$|]						*	*	*	
1	/Q-S/ → [|Q^R $S_i^R S_i$|]						*	*		

Output of EVAL

/Q-S/ → \|Q^R $S_i^R S_i$\|

/Q-S-S/ → \|Q^R $S_i^R S_i$\| MAX-C

In the second mapping /Q-S/ → [\|Q^R $S_i^R S_i$\|] the input /S/ spreads, earning this mapping a violation of INTEGRITY. The two surface [S]'s are both affiliated with the stem, so that this mapping also violates OCP_{Stem}. In the first mapping /Q-S-S/ → [\|Q^R $S_i^R S_i$\|] the second input /S/ deletes, earning this mapping a violation of MAX-C. The remaining /S/ then spreads so that this mapping also violates INTEGRITY. In this mapping the two surface [S]'s are both affiliated with the stem, so that this mapping also violates OCP_{Stem}. The extra MAX-C violation of the first mapping implies that the second mapping is the more harmonic of the two. This means that upon hearing a word such as [QiSeS], a Hebrew speaker will infer that this word was derived from a bi-consonantal rather than a tri-consonantal root.

The grammar of Hebrew can successfully handle roots with identical contiguous consonants (see the tableau in (23) above). However, due to lexicon optimization Hebrew speakers will never infer the existence of such roots.

2.1.7 The processing of non-words

Now that we have a mini grammar for Hebrew, we can consider how Hebrew speakers will process non-words. In the experiments that I will discuss below (§2.2), Berent *et al.* used three kinds of non-words, namely non-words with three non-identical consonants in the stem positions ([|QiSeM|]), non-words with identical consonants in the last two stem positions ([|QiSeS|]), and non-words with identical consonants in the first two stem positions ([|QiQeS|]).

I will assume that for any geminate non-word percept, listeners have two choices for what the input root for the percept could be. They can assume that the input root is identical to the consonants of the stem in the perceived non-word, or they can assume that the input root is bi-consonantal. For a percept [|QiSeS|] they can therefore assume that the input form of the root is /Q-S-S/ or that it is /Q-S/. For a percept [|QiQeS|] listeners can assume that input form of the root is /Q-Q-S/ or that it is /Q-S/. However, for a non-word percept with three non-identical consonants in stem position, listeners will consider only one possible input for the root, namely a form that is identical to the consonants in the stem of the non-word percept. For a [|QiSeM|] percept, listeners will only consider /Q-S-M/ as input for the root.

(25) **Underlying representation for roots in non-words**

	Percept	**UR considered**
No-geminate:	[QiSeM]	/Q-S-M/
Final-geminate:	[QiSeS]	/Q-S/ and /Q-S-S/
Initial-geminate:	[QiQeS]	/Q-S/ and /Q-Q-S/

The tableau below in (26) compares each of the mappings that listeners will consider when they are processing non-words.

(26) **Comparing non-words**

Geminate?		ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$CONTIG_{Root}$
No	1 /Q-S-M/ → $[\|Q^R S^R M^R\|]$									
Final	2 /Q-S/ → $[\|Q^R S_i{}^R S_i\|]$						*	*		
	3 /Q-S-S/ → $[\|Q^R S^R S^R\|]$				*		*			
Initial	4 /Q-S/ → $[\|Q_i Q_i{}^R S^R\|]$	*					*	*		
	3 /Q-Q-S/ → $[\|Q^R Q^R S^R\|]$				*		*			

Output of EVAL

Cut-off

/Q-S-M/ → $[\|Q^R S^R M^R\|]$ No-gemination

|

/Q-S/ → $[\|Q^R S_i{}^R S_i\|]$ Final-gemination

|

/Q-S-S/ → $[\|Q^R S^R S^R\|]$ Final/Initial-gemination

/Q-Q-S/ → $[\|Q^R Q^R S^R\|]$

|

/Q-S/ → $[\|Q_i Q_i{}^R S^R\|]$ Initial-gemination

This comparison shows the following: (i) Non-words with no-gemination are the most well-formed. (ii) Depending on what underlying representation is assumed, final-gemination non-words are either more well-formed than initial-gemination non-words or

equally well-formed with initial-gemination non-words. (iii) Initial-gemination non-words are either less well-formed than final-gemination non-words or equally well-formed with final-gemination non-words.

Suppose that every time a listener is presented with a non-word with gemination, he/she considers both potential input forms and selects the form that results in the most harmonic parsing (i.e. performs lexicon optimization every time). The listener will then eventually settle on bi-consonantal input roots for final-geminate forms. The derivation from a bi-consonantal root to final-geminate is more well-formed than the derivation of an initial-geminate form, irrespective of whether the initial-geminate is derived from a bi-consonantal or a tri-consonantal root. The prediction is therefore that final-geminates will be more well-formed than initial-geminates.

(27) **Well-formedness of different kinds of non-words**

No-gemination	[QiSeM]	Most well-formed
	|	
Final-gemination	[QiSeS]	
	|	
Initial-gemination	[QiQeS]	Least well-formed

If language users use the grammar in the processing of non-words, then we would expect to see effects of this relation in the way in which they process non-words. In particular we would expect that: (i) In well-formedness judgment experiments language users would rate no-gemination non-words better than final-gemination non-words, and final-gemination non-words again better than initial-gemination non-words. (ii) In lexical decision tasks, language users should on average reject initial-gemination non-words more quickly than final-gemination non-words, and final-gemination non-words again more quickly that no-gemination non-words. These predictions are represented

graphically in (28). In the next section I discuss a series of experiments performed by Berent *et al*. in which (most of) these predictions are confirmed.

(28) **Predictions with regard to well-formedness judgments and lexical decision**

Well-formedness judgments	**Rank-ordering imposed by EVAL**	**Lexical decision RT's**
Decreasing well-formedness judgments ↓	No-gemination	Decreasing reaction time in lexical decision ↓
	|	
	Final-gemination	
	|	
	Initial-gemination	

2.2 Word-likeness and lexical decision in Hebrew

Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997) performed a series of experiments in which they tested whether the restriction on the distribution of identical consonants in Hebrew roots influences the manner in which speakers of Modern Hebrew process non-words. These experiments included both well-formedness judgment experiments and lexical decision experiments. In the well-formedness judgment experiments, the subjects reacted according to the predictions of the analysis developed above – that is, no-gemination rated best, then final-gemination and then initial-gemination. The results of the lexical decision experiments agree only partially with the predictions. In accordance with the predictions, final-geminate forms were rejected slower than initial-geminate forms. However, contra the predictions, final-geminate forms were also rejected slower than no-geminate forms. (See §2.2.2 below for a discussion of what could have caused the slowdown in reaction times associated with geminate forms.)

In the rest of this section I will first discuss the results of the well-formedness judgment experiments performed by Berent *et al*., and then the results of their lexical decision experiments. I will not discuss their experimental design in detail, for the following two reasons: (i) The information about their experimental design is discussed in detail in the published reports of their research. (ii) Their experiments were designed to test hypotheses that are different from the hypotheses that I discuss here. Many of the details of their experimental design are therefore not relevant to the discussion here.

2.2.1 Well-formedness judgment experiments

Berent *et al*. conducted two kinds of well-formedness judgment experiments, namely gradient well-formedness judgment experiments and comparative well-formedness judgment experiments. In the gradient well-formedness judgment experiments subjects are presented with individual tokens and are required to rate each token on a 5-point scale for its well-formedness. In the comparative well-formedness judgment experiments subjects are presented with three tokens at a time, and they are required to arrange the three tokens in the order of their well-formedness.

2.2.1.1 Gradient well-formedness judgment experiments

There are three separate studies in which Berent *et al*. conducted gradient well-formedness judgment experiments. These are Berent *et al*. (2001a), Berent *et al*. (2002), and Berent and Shimron (1997). There are slight differences in design between the three experiments. However, these differences are small and not relevant to the point of the current discussion. The results in all three of experiments were basically the same. I will therefore discuss one of the three experiments as a representative example. For this purpose I select the results of Berent and Shimron (1997).

Berent and Shimron selected 24 tri-consonantal roots, each consisting of three non-identical consonants. They also selected 24 bi-consonantal roots, each consisting of two non-identical consonants. None of these 48 roots corresponded to actual roots of Hebrew. They conjugated each of the 24 tri-consonantal roots in three different verbal forms, so that there were 72 non-words formed from tri-consonantal roots. The 24 bi-consonantal roots were transformed into initial and a final-geminate stems by doubling either the initial or final consonant of each root. Each of these geminates was then also conjugated in three verbal forms, so that there were 72 initial-geminate non-words and 72 final-geminate non-words. In total, the stimulus list for this experiment consisted of 216 non-words conjugated as verbs.

This list of forms was randomized and presented in written form to 15 native speakers of Hebrew. The task of the subjects was to rate each non-word as to its word-likeness or well-formedness. Forms were rated on a 5-point scale, where [5] indicated a form that sounded excellent as a possible word, and [1] a form that sounded impossible as a word of Hebrew. Based on the analysis developed above (§2.1), the expectation is that no-geminate forms would on average be rated better than final-geminate forms, which again would be rated better than initial-geminate forms – the less marked a form, the better it would be rated.

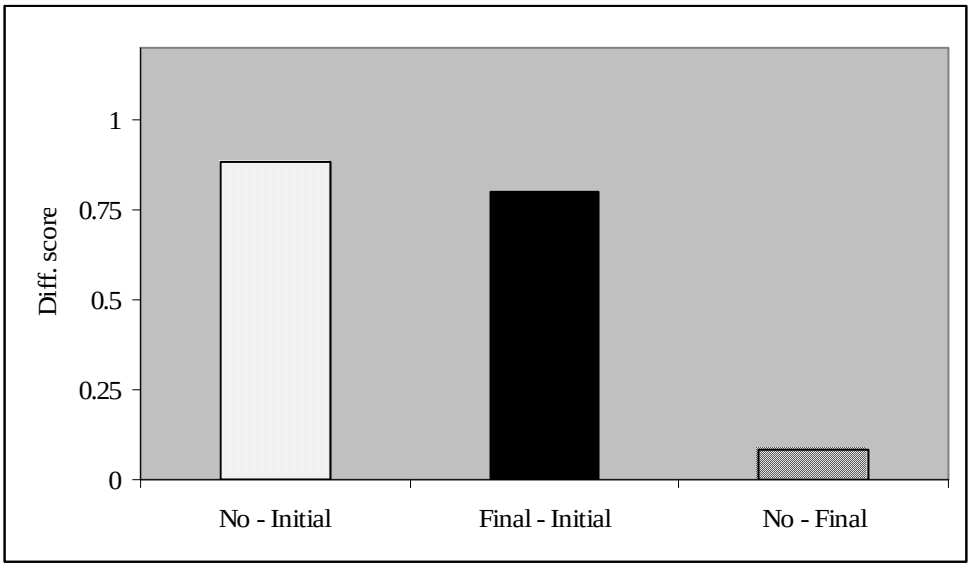
Berent and Shimron do not report the average scores assigned to each of the three token types. However, they do report the difference scores – i.e. the difference between

the average ratings assigned to each of the three token types.[24] The results of this experiment are summarized in the table in (29), and represented graphically in (30).

(29) **Average well-formedness ratings in the gradient well-formedness judgment experiment of Berent and Shimron (1997)**

Comparison	**Example**	**Difference score**	***t***	***df***	***p***
Initial-gemination and no-gemination	Q-Q-S Q-S-M	0.881	11.139	46	< 0.000
Initial-gemination and final-gemination	Q-Q-S Q-S-S	0.801	9.984	46	< 0.000
Final-gemination and no-gemination	Q-S-S Q-S-M	0.081	_[25]		> 0.05

(30) **Gradient well-formedness ratings: Difference scores between the mean ratings for the different token classes in the Berent and Shimron (1997) data**



[24] The difference scores were computed by subtracting the average score for the more marked token type from the average score for the less marked token type – i.e. a positive difference score means that the less marked token type was on average rated better. Specifically: (i) Difference Score (Initial~Final) = Mean Score (Final) – Mean Score (Initial). (ii) Difference Score (Initial~No) = Mean Score (No) – Mean Score (Initial). (iii) Difference Score (Final~No) = Mean Score (No) – Mean Score (Final).

[25] Berent and Shimron unfortunately do not report the *t*-statistic for this comparison. They do, however, report that a *p*-value of larger than 0.05 was obtained for this comparison using the Tukey HSD test – i.e. there was no significant difference between the mean ratings for these two classes of tokens.

In this experiment there is evidence for the fact that initial-geminate forms were rated worse than both final-geminate forms and no-geminate forms. Since forms with initial-gemination are predicted to be the most marked by the analysis developed above, this corresponds to our expectation. But what about the comparison between no-geminate and final-geminate forms? Although both of these classes represent possible words of Hebrew, under the analysis developed above final-geminate forms are more marked than no-geminate forms. The prediction is that final-geminate forms should receive lower scores than no-geminate forms. In absolute terms this was confirmed by the results of the experiment – no-geminate forms were rated on average 0.081 points better than final-geminate forms. However, this is a very small difference and it did not reach the critical level of statistical significance.

What should we make of the failure to find a difference between no-geminate and final-geminate forms? First, it should be noted that the results do not go against the predictions of the analysis developed above in §2.1 – it simply does not confirm these predictions. Secondly, Berent and Shimron claim that the non-result in this condition can be attributed to the experimental design. Both no-geminate and final-geminate forms are possible words of Hebrew. It is then possible that both of these types of words were rated very well, and that there was simply not enough room at the top end of the 5-point scale to differentiate between these two classes of tokens. The non-difference between these two token types can therefore be attributed to a ceiling effect.[26]

---

[26] In order to counter this problem, Berent and Shimron conducted a second type of well-formedness judgment experiment in which subjects had to rate final-geminates and no-geminates in relation to each other – i.e. the experimental design forced a direct comparison between these two kinds of forms. In this comparative well-formedness experiment, they did find evidence for the predicted difference between final-geminate forms and no-geminate forms. See §2.2.1.2 just below.

2.2.1.2 Comparative well-formedness judgment experiments

Berent and Shimron (1997) and Berent *et al.* (2001a, 2002) also conducted comparative well-formedness judgment experiments. Since the results of the three replications of these experiments were basically the same, I discuss only the results of Berent and Shimron (1997) here are a representative example.

In the comparative well-formedness judgment experiment, Berent and Shimron used the same 48 roots (24 tri-consonantal and 24 bi-consonantal) as in the gradient well-formedness judgment experiment. They matched each bi-consonantal root with a tri-consonantal root that differed from the bi-consonantal root by the addition of one consonant. Each pair of bi- and tri-consonantal roots was converted into a triple of tri-consonantal stems by geminating the first and last consonant of the bi-consonantal root. A stem triple consisted of a stem with three non-identical consonants, an initial-geminate and a final-geminate. The tri-consonantal stem shared two of its consonants with the two geminate stems. These triples were conjugated in three different verbal forms of Hebrew, so that each triple corresponded to three triples of non-words conjugated as verbs. A stem triple |Q-S-M|~|Q-S-S|~|Q-Q-S| could, for in stance, correspond to the following three non-word triples: (i) [|QiSeM|]~[|QiSeS|]~[|QiQeS|], (ii) [|maQSiMim|]~[|maQSiSim|]~[|maQQiSim|], (iii) [|hitQaSeMtem|]~[|hitQaSeStem|]~[|hitQaQeStem|]. Each of the 24 triples corresponded to three such non-word triples, for a total of 72 non-word triples.

These 72 non-word triples were randomized and presented to 18 native speakers of Hebrew in written form. Subjects had to order the members of each triple according to their acceptability as words of Hebrew. A score of [3] was assigned to the member that was considered to be most word-like, a score of [1] to the member that was considered

least word-like, and a score of [2] to the remaining member. This experimental design results in a direct comparison between no-gemination, final-gemination and initial-gemination forms. Based on the analysis developed above (§2.1), we expect no-geminate forms to be rated better than final-geminate forms, which again would be rated better than initial-geminate forms – the less marked a form, the better it would be rated.

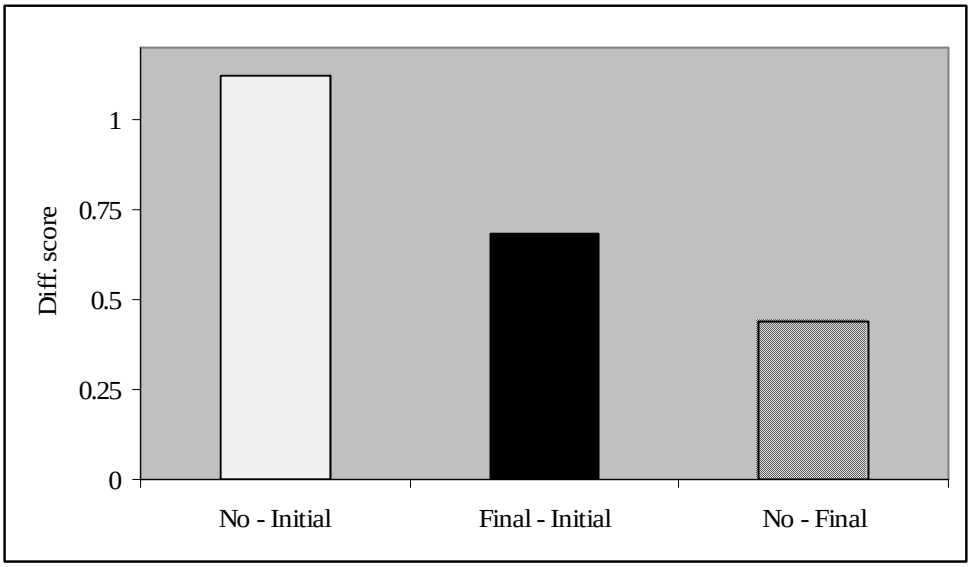
As in the gradient well-formed judgment experiment, Berent and Shimron do not report the average scores assigned to each of the three token types, but only the difference scores between the three token types. The results of this experiment are summarized in the table in (31), and represented graphically in (32).

In accordance with the predictions of the analysis in §2.1, the initial-geminate forms were rated worse than both the final-geminate forms and the no-geminate forms. This corresponds to the fact that initial-geminate forms are predicted to be more marked than both final-geminate forms and no-geminate forms. In the gradient well-formedness judgment experiment no difference was found between the two kinds of possible words, i.e. between final-geminate forms and no-geminate forms. That null result was attributed to the experimental design – no direct comparison was possible between these two kinds of tokens and they received equally good ratings as possible words. In the comparative well-formedness experiment, subjects are required by the experimental design to compare these two forms directly. If there is a difference between these two kinds of tokens, we expect to find evidence of that difference in this experiment. And no-geminate forms were indeed preferred over final-geminate forms in this experiment. Even though both final-geminate and no-geminate forms are possible words of Hebrew, Hebrew speakers consider no-geminate forms as more word-like/well-formed than final-geminate forms.

(31) **Average well-formedness ratings in the comparative well-formedness judgment experiment of Berent and Shimron (1997)**

Comparison	Example	Difference score	*t*	*df*	*p*
Initial-gemination and no-gemination	Q-Q-S Q-S-M	1.122	18.55	46	< 0.000
Initial-gemination and final-gemination	Q-Q-S Q-S-S	0.682	11.28	46	< 0.000
Final-gemination and no-gemination	Q-S-S Q-S-M	0.44	_[27]		< 0.05

(32) **Comparative well-formedness: Difference scores between the mean ratings for the different token classes in the Berent and Shimron (1997) data**



This experiment confirms the predictions of the analysis developed in §2.1. In terms of well-formedness judgment experiments we have no evidence against the predictions of the analysis, and some evidence in favor of this analysis. Consequently, I interpret the results of the well-formedness judgment experiments as confirming the predictions of the analysis developed in §2.1 above. The less marked a token is according to the grammar, the more well-formed/word-like it is judged to be by language users.

[27] Berent and Shimron do not report the *t*-score for this comparison. They do report that a *p*-value of smaller than 0.05 was obtained for this comparison using the Tukey HSD test.

2.2.2 Lexical decision experiments

Both Berent *et al*. (2001b) and Berent *et al*. (2002) performed experiments to determine whether the restriction on the distribution of identical consonants influences the manner in which Hebrew speakers perform lexical decision tasks. The results of all of these experiments were basically the same, and I will therefore discuss one of the experiments as representative. For this purpose I select *Experiment 2* of Berent *et al*. (2001b).

Berent *et al*. selected 30 tri-consonantal roots, each consisting of three non-identical consonants. They also selected 30 bi-consonantal roots, each consisting of non-identical consonants. Not one of these roots corresponded to actual roots of Hebrew. They transformed each of the bi-consonantal roots into two tri-consonantal stems by geminating either the first or the last consonant of the root. This resulted in 90 stems. Of these 90 stems, 30 had no identical consonants (no-gemination), 30 had identical consonants in the initial two positions (initial-gemination), and 30 had identical consonants in the final two positions (final-gemination). Each of these stems was inserted into the same verbal pattern of Hebrew, resulting in 90 non-words, conjugated as verbs.

Berent *et al.* also selected 45 tri-consonantal and 45 bi-consonantal Hebrew roots. These roots all corresponded to actual Hebrew verbal roots. These 90 roots were inserted into the same verbal forms as the non-words, resulting in 90 actual Hebrew roots. The 90 words and 90 non-words were randomized into one list. The list was presented in a lexical decision task to 20 native speakers of Hebrew. The tokens were presented one at a time on a computer monitor. Subjects had to indicate whether the token was or was not a word of Hebrew by pushing one of two buttons. Response times were recorded. Statistical analyses were done on the response times to correct non-word responses.

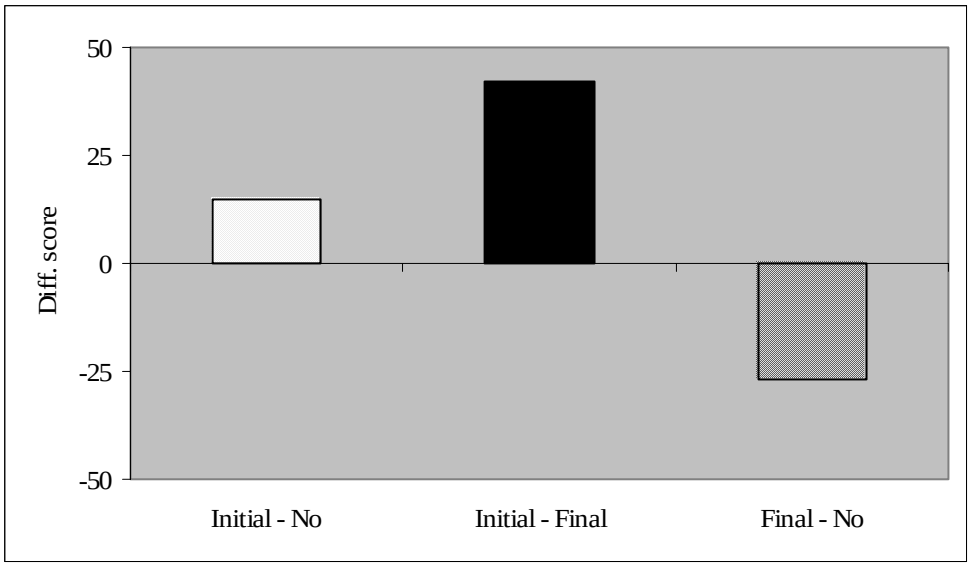
Recall the basic prediction with regard to lexical decision – the more well-formed a non-word token is according to the grammar, the more seriously the language user will consider it as an actual word of Hebrew. Consequently, we expect slower reaction times with more well-formed non-words than with less well-formed non-words. In terms of the analysis developed above in §2.1, we therefore expect that initial-geminate forms should be rejected the most quickly, final-geminate forms more slowly, and no-geminate forms the most slowly. The results of this experiment confirmed some of these predictions. In particular, the results confirmed the prediction that initial-geminate forms should be rejected more quickly than final-geminate forms. However, final-geminate forms were not rejected more quickly than no-geminate forms. In fact, the opposite was found – final-geminate forms were rejected more slowly than no-geminate forms. The table in (33) summarizes the results, and (34) contains a graphic representation of the results.

In the analysis developed in §2.1, final-geminate forms are less well-formed than no-geminate forms. Under the assumption that language users use the information provided by grammar when they make lexical decisions, the prediction was that final-gemination non-words should be responded to more quickly than no-gemination non-words. However, the results of the experiment counter this – final-geminate forms were rejected more slowly than no-geminate forms. Similarly, the analysis developed in §2.1 implies that initial-geminate forms are more marked than no-geminate forms. The expectation is therefore that initial-geminate forms should be reacted to more quickly than no-geminate forms. This was also not confirmed in the experiments – no significant difference was found in the reaction times associated with initial-geminate forms and no-geminate forms. How should we interpret these negative and null results?

(33) **Average difference in response times between different types of non-words in *Experiment 2* of Berent *et al.* (2001b)**

Comparison	Example	Difference score (*ms*) [28]	*t*	*df*	*p*
Initial-gemination and no-gemination	Q-Q-S Q-S-M	15	–	–	> 0.05[29]
Initial-gemination and final-gemination	Q-Q-S Q-S-S	42	4.09	56	< 0.0001[30]
Final-gemination and no-gemination	Q-S-S Q-S-M	-27	2.48	56	< 0.009[31]

(34) **Lexical decision: Difference scores between the mean response times for the different token classes in the Berent *et al*. (2001b) data**



[28] These difference scores were computed by subtracting the average response time of the more marked token types from the average response times of the less marked token types. A positive difference score then indicates that the more marked tokens were rejected more quickly, and therefore counts as a success in terms of the hypothesis.

[29] Berent *et al*. do not report the actual *t*-statistic for this comparison. They do, however, state that there was no significant difference in mean response times for the initial-geminate forms and the no-geminate forms (Berent *et al*., 2001b:655).

[30] This *p*-value is for a one-tailed *t*-test, assuming a positive difference score. It therefore shows that initial-geminate forms were reacted to significantly more quickly than final-geminate forms.

[31] This *p*-value is for a one-tailed *t*-test, assuming a negative difference score. It therefore shows that final-geminate forms were reacted to significantly more slowly than no-geminate forms.

Berent *et al*. explain these unexpected results as follows: The more productive (phonological or morphological) processes are applied in the formation of a non-word, the more word-like it will be experienced to be by language users. An example from English might make the idea clearer: An English non-word such as “blick” will be experienced as less word-like than a non-word such as “blicked”, because a productive derivational process has applied in “blicked”. Berent *et al*. then assume that the subjects in their experiment interpreted all geminate forms (both initial and final) as derived via a productive gemination process from bi-radical roots. The fact that gemination was applied in the formation of these non-words lends a word-like feeling to them. This is then supposed to explain the slower than expected response times associated with geminate non-words.

However, there is a problem with this explanation of Berent *et al*. If application of process results in causing a non-word to be experienced as more word-like, then the effects of this should have been seen also in the word-likeness/well-formedness judgment experiments. For instance, we would then have expected that final-geminate forms would have been rated as more word-like than no-geminate forms. And this was not found. In fact, the opposite was found. I therefore suggest a different explanation for the unexpected response time results associated with geminate forms.

In no-geminate forms, the subjects consider only one possible derivation for the percept – from a tri-consonantal root to a tri-consonantal stem. However, both in initial and final-geminate forms, subjects consider two possible derivations – from a bi-consonantal root to a tri-consonantal stem, and from a tri-consonantal root to a tri-consonantal stem. (See the discussion in §2.1.7 about this.) Two derivational histories

have to be considered for the geminate forms, and this is what slows down the reaction times associated with geminate forms.

Consider the comparison between initial-geminate and final-geminate forms. For both of these two kinds of tokens, subjects have to consider two derivational histories. There is therefore no processing difference between these two kinds of tokens. The only difference between them is in terms of their markedness/well-formedness. We do therefore expect to see the result predicted by the grammar – i.e. that initial-geminate forms should be rejected more quickly than final-geminate forms. And this is confirmed by the results of the experiment.

Now consider the comparison between the no-geminate forms and the two types of geminate forms (final and initial). According to the analysis developed in §2.1 these forms are related as follows in terms of their well-formedness: |No-gemination ™ Final-gemination ™ Initial-gemination|. Based on grammar alone, we would expect the following relation between the reaction times: RT(No-gemination) > RT(Final-gemination) > RT(Initial-gemination). Suppose that the additional processing associated with gemination adds about the same amount of time to both final and initial-geminate forms. We can then explain the results of the experiment as follows: (i) Based on grammar alone, we expect a smaller RT-difference between no-geminate forms and final-geminate forms, than between no-geminate forms and initial-geminate forms. (ii) The additional processing time added by gemination is longer than the expected RT-difference between no-gemination and final-geminate forms. The result is that final-geminate forms are rejected more slowly than no-geminate forms. (iii) However, the additional time added by the gemination is about as long as the expected RT-difference

between no-geminate and initial-geminate forms. The result is that there is no appreciable difference in actual RT's between initial-geminate and no-geminate forms.

Under this interpretation we can also explain why the well-formedness judgment experiments did give the results predicted by the grammatical analysis. Also in the well-formedness judgment experiments subjects consider two possible derivational histories for the two kinds of geminate forms. However, these experiments were all self paced – i.e. subjects could take as much time as they needed to respond. The additional processing time in the geminate forms should therefore not influence the judgments of the subjects.

3. The processing of non-words of the form [sCvC] in English

In the previous section we saw evidence for the fact that language users make multi-level well-formedness distinctions. Hebrew speakers distinguish not only between possible and impossible words. They also distinguish between different possible words in terms of their well-formedness. This shows that we need a theory of grammar that can make multi-level well-formedness distinctions. In this section I will discuss a restriction on possible words in English that confirms this. I will discuss a set of experiments that show that English speakers distinguish between possible and impossible words in terms well-formedness. However, the results of the experiments show that in addition to this distinction English speakers also distinguish between different kinds of impossible words in terms of well-formedness. Together with the Hebrew results in §2, these results show that we need a theory of grammar that can make the following distinctions: (i) between possible and impossible words; (ii) within the set of possible words, between more and less well-formed possible words; (iii) also within the set of impossible words, between

the more and less well-formed. We need a theory of grammar that can take any two forms and compare them for their relative well-formedness, whether they are possible or impossible words.

English places restrictions on what consonants can co-occur in the onset and coda position of a mono-syllabic word.[32] Restrictions of this kind were first noted by Fudge (1969), and include for example the following: (i) A mono-syllabic word cannot begin with [Cr] and end in [r] – i.e. **frer*, **krer*, etc. (ii) A mono-syllabic word cannot begin in [s] plus a nasal and end in nasal – i.e. **snam*, **smang*, etc. (iii) A mono-syllabic word cannot begin in [s] plus a nasal and end in [lC] – i.e. **snelk*, **smelk*, etc.

In this section I will investigate one of these restrictions in more detail. English does not allow words of the form [sCvC] where both C's are voiceless labial stops or voiceless velar stops. [sCvC]-words are allowed with two voiceless coronal stops – i.e. *state* is a word of English, but **spape* and **skake* are not even possible words (Browne, 1981, Clements and Keyser, 1983, Davis, 1982, 1984, 1988a, 1988b, 1989, 1991, Fudge, 1969, Lamontagne, 1993: Chapter 6).[33]

---

[32] These restrictions actually apply to syllables. A multi-syllabic word that has a syllable with any of these combinations will also not be a possible word of English. In the discussion here I will limit myself to mono-syllabic forms so that restrictions on possible syllables can also be treated as restrictions on possible words.

[33] This is also part of a larger restriction. For instance, a form with a voiced and voiceless labial is also not allowed, i.e. **spab*. However, forms with voiced and voiceless velars are at least marginally tolerated, cf. *skag*. Forms with voiceless stops and homorganic nasals are also generally not allowed, i.e. **skang*, **spim*.

The term *spam* is an exception. This term originated as a brand name for a kind of luncheon meat *SPAM*, and was later (inspired by a Monty Python sketch) extended to refer to unsolicited e-mail (http://en.wikipedia.org/wiki/Spamming#Etymology). Brand names, like other proper names, are often exempted from restrictions that apply to other words.

In the discussion here I will focus on this restriction only as it applies to the voiceless stops.

In the rest of this section I will discuss experimental evidence showing that this restriction has psychological reality for speakers of English. In particular, it influences the way in which they process non-words. This section is structured as follows: In §3.1 I develop an OT analysis of this restriction on possible words in English. The analysis establishes that [sCvC]-forms are related as follows in terms of their relative well-formedness: |sTvT ™ sKvK ™ sPvP|. In §3.2 I then discuss a set of well-formedness judgment and lexical decision experiments that I conducted to test the predictions of this analysis. The results of the experiments confirm that the restriction on [sCvC]-forms are psychologically real for speakers of English and that it influences the way in which they process non-words of this form.

3.1 Restrictions on [sCvC]-words

English does not tolerate words of the form [sCvC] where both C's are voiceless labial or velar stops. Forms with two voiceless coronal stops or with two heterorganic voiceless stops are tolerated.

(35) **Restrictions on [sCvC]-forms in English**

Allowed			**Not allowed**	
Two [t]'s		*state*	Two [p]'s	**spape*
Different C's	[k]/[t]	*skate, steak*	Two [k]'s	**skake*
	[p]/[t]	*spit, steep*		
	[p]/[k]	*skip, speak*		

The same restriction is found in German. Twaddell (1939, 1940) compiled a list of all the mono- and bi-syllabic words in Duden's 1936 German dictionary. This resulted

in a list of approximately 37,500 tokens. He then calculated the frequency of all [(C)(C)(C)v(C)(C)(C)]-sequences (where [v] is a stressed vowel).[34] Twaddell found many words with the sequence [sTvT]. However, he found no words with the sequence [sKvK] or [sPvP]. The data in the table in (36) is extracted from Twaddell's 1940-paper.

(36) **Frequency count of [sCvC] sequences in German**[35]

st			**sk**			**sp**		
t	k	p	t	k	p	t	k	p
51	95	76	15	–	2	58	73	–

Like English, German tolerates [sCvC] words where the two C's are heterorganic voiceless stops, or two voiceless coronal stops, but not where C's are voiceless labial or velar stops. The examples in (37) illustrate this point.[36]

Afrikaans is different from German and English. Like German and English Afrikaans does not tolerate [sCvC] words where both C's are voiceless labial stops. However, unlike German and English, forms with two voiceless velar stops are well-formed in Afrikaans. This is illustrated by the examples in (38).[37]

---

[34] Twaddell did not take syllabic boundaries into consideration. His counts would therefore include also forms with syllable boundaries after the vowel, i.e. [sCv.Cv…]. For the purposes of the discussion here, what is important is that syllables of the form [sCvC] are included in his counts. Another caveat is in order – Twaddell did not include the unstressed syllables in his counts. His counts alone are therefore not enough to substantiate the claim that German has the same restriction as English. The intuition that [sPvP] and [sKvK] sequences are ill-formed in German was confirmed by consultation with native speakers of German. See footnote 38 below.

[35] Since [s] preceding a consonant is often pronounced as [ʃ] in German, Twaddell included both [sC] and [ʃC] sequences in these counts.

[36] I am indebted to my German consultants, Florian Schwarz and Tanja Vignjevic, for supplying these examples. They also confirmed the intuition that [sPvP] and [sKvK] forms are ill-formed in German.

[37] Afrikaans is my native language. I therefore base my statements about Afrikaans on my own intuitions.

(37) **Restrictions on [sCvC]-forms in German**

Allowed				**Not allowed**	
Two [t]'s		*Staat*	"state"	Two [p]'s	**Spaap*
		Stadt	"city"		**Spep*
Different C's	[k]/[t]	*Stock*	"stick"	Two [k]'s	**Skaak*
		Skat	"card game"		**Skek*
	[p]/[t]	*Stop*	"stop"		
		Spott	"ridicule"		
	[p]/[k]	*Skip*	"leader of team in curling"		
		Speck	"bacon"		

(38) **Restrictions on [sCvC]-forms in Afrikaans**

Allowed				**Not allowed**	
Two [t]'s		*staat*	"state"	Two [p]'s	**spaap*
		stad[38]	"city"		**spep*
Two [k]'s		*skok*	"shock"		
		skaak	"chess"		
Different C's	[k]/[t]	*stok*	"stick"		
		skat	"treasure"		
	[p]/[t]	*stop*	"stop"		
		spot	"ridicule"		
	[p]/[k]	*skip*	"ship"		
		spek	"bacon"		

[38] Afrikaans has coda devoicing. This word therefore ends on a [t] in pronunciation.

[sTvT]-forms are possible words in English, German and Afrikaans, and [sPvP]-forms are not possible words in any of these languages. [sKvK]-forms take the intermediate position. These forms are possible words in Afrikaans, but not in English or German. Based on this, I propose the following markedness/harmony scale: |sTvT ™ sKvK ™ sPvP|.

In the rest of this section I develop an OT analysis of this restriction on possible words in English. In this analysis I will assume the markedness/harmony scale |sTvT ™ sKvK ™ sPvP|. Even though neither [sKvK]-forms nor [sPvP]-forms are possible words in English, I will assume that [sKvK]-forms are more well-formed than [sPvP]-forms. Section §3.1.1 contains a basic OT analysis. Section §3.1.2 then contains an extensive motivation of the analysis, showing that it is well founded both cross-linguistically and theory internally. Section §3.1.3 considers the predictions that follows from the analysis with regard well-formedness judgments and lexical decision reaction times.

3.1.1 A basic OT analysis

Following a tradition that originated with Prince and Smolensky (1993), I assume that there is a markedness constraint against every element on a harmony scale. The markedness constraints are ranked in an order opposite to the elements on the harmony scale, so that the element that occupies the lowest slot on the harmony scale violates the highest ranked markedness constraint. This is illustrated in (39) for the harmony scale associated with [sCvC]-forms. [39]

---

[39] De Lacy (2002) has a different view. Rather that assuming a fixed ranking between markedness constraints that refer to a harmony scale, he argues for constraints that can freely rerank but that are in a stringency relation (Prince, 1998). Under this interpretation of the relationship between harmony scales and markedness constraints [sTvT] will violate only *sTvT, [sKvK] will violate *sKvK and

(39) **From harmony scales to markedness constraint hierarchy**

Harmony scale: |sTvT ™ sKvK ™ sPvP|

Markedness constraint hierarchy: ||*sPvP o *sKvK o *sTvT||

(Where ***sC$_i$vC$_i$** = Do not allow the sequence [sC$_i$vC$_i$] within a single syllable.)

English tolerates violation of the constraint *sTvT. This implies that all faithfulness constraints that could be violated in order to avoid violation of *sTvT should outrank *sTvT. On the other hand, English does not tolerate violation of the constraints *sKvK or *sPvP. These two constraints must therefore outrank at least some faithfulness constraint that could be violated in order to avoid violation of *sKvK or *sPvP. Since there is no active process in English that shows how English would avoid violation of *sPvP and *sKvK, it is not possible to decide definitively what the relevant faithfulness constraint is. Violation of these constraints can be avoided in many different ways. In (40) some possibilities are listed, together with the faithfulness constraint that each repair strategy would violate.

---

*sTvT, and [sPvP] will violate *sPvP, *sKvK, and *sTvT. Although the constraints can freely rerank, the harmonic ordering |sTvT ™ sKvK ™ sPvP| will still hold. Whether we use markedness constraints in a fixed ranking or stringency related constraints, the same harmonic ordering between the [sCvC]-forms hold. The main claims that I make in this section is that this is the harmonic ordering that grammar imposes on [sCvC]-tokens and that this ordering influences the manner in which English listeners process [sCvC]-tokens. Since this ordering is imposed both by constraints in a fixed ranking and by stringency related constraints, it does not matter for my purposes which of the two analyses I use.

Gouskova (2003) uses constraints in a fixed ranking, but shows that there is no constraint against the most harmonic element on a harmony scale. However, even under Gouskova's interpretation of harmony scales there will be a constraint against [sTvT]-forms. The reason is that [sTvT] is not the least marked form. Forms with heterorganic voiceless stops probably occur at the least marked end of the harmonic ordering, i.e. |sTvK ™ sTvT ™ sKvK ™ sPvP|. The existence of the constraint *sTvT is then not contrary to Gouskova's interpretation of harmony scales.

(40) **Avoiding violation of *sCvC**

Input	Repair	Example output	Faithfulness constraint violated
/sKvK/	Deletion	[sKv], [KvK]	MAX
	Epenthesis	[sKv.Kv]	DEP
	Place change	[sKvT], [sPvK]	IDENT[place]

In the rest of the discussion I will assume that the relevant faithfulness constraint is IDENT[place]. However, this is an arbitrary choice – any of the other constraints in (40) would have done equally well. If we rank IDENT[place] between *sKvK and *sTvT, we can account for the fact that [sTvT] is a possible word of English, while [sKvK] and [sPvP] are not. This is shown in the tableaux in (41).

(41) **a. [sTvT] possible word**

/sTvT/ → [sTvT]

/sTvT/	*sPvP	*sKvK	IDENT[place]	*sTvT
sTvT				*
sTvK			*!	

b. [sKvK] not possible word

/sKvK/ → [sKvT]

/sKvK/	*sPvP	*sKvK	IDENT[place]	*sTvT
sKvK		*!		
sKvT			*	

c. [sPvP] not possible word

/sPvP/ → [sPvT]

/sPvP/	*sPvP	*sKvK	IDENT[place]	*sTvT
sPvP	*!			
sPvT			*	

These tableaux show that an /sTvT/-input will be mapped faithfully onto itself, confirming that [sTvT] is a possible word. However, for an /sKvK/-input, an unfaithful candidate will be selected over the faithful [sKvK]. This shows that [sKvK] is not a possible word of English. The same is true for an /sPvP/-input.

3.1.1.1 The critical cut-off

We need to locate the critical cut-off point in this mini-hierarchy for English. (For a discussion of the critical cut-off, refer to Chapter 1 §2.2.3, Chapter 3 §2.3.) Due to the lack of evidence of variation in the pronunciation [sCvC]-forms in English, I make the conservative assumption that the critical cut-off is located at the bottom of the mini-hierarchy developed above (see Chapter 4 §3.1 on this assumption). All candidates will therefore violate at least some constraint ranked higher than the cut-off. In these situations, only the best candidate is accessible as potential output. This assures that no variation will be observed in the pronunciation of [sCvC]-forms in English. This is illustrated for an /sTvT/-input in (42). However, the same point can be made for /sPvP/-inputs and /sKvK/-inputs. As always, a thick vertical line is used to indicate the place of the critical cut-off in the hierarchy, and a graphic representation of the ordering imposed by EVAL on candidate set is given next to the tableau.

(42) **Adding in the critical cut-off**

/sTvT/ → [sTvT]

/sTvT/	*sPvP	*sKvK	IDENT[place]	*sTvT
1 sTvT				*
2 sTvK			*!	

Output of EVAL

Cut-off

sTvT *sTvT

sTvK IDENT[place]

3.1.1.2 Interim summary

The table in (43) contains a summary of the constraints and rankings argued for thus for. The first column contains the rankings, the second column a short motivation for the ranking, and the third column an indication of where in the preceding discussion the ranking was discussed.

(43) **Summary of the constraints and rankings for English**

Ranking	**Motivation**	**Where motivated**
*sTvT o {*sKvK, *sPvP}	[sTvT] possible word [sKvK], [sPvP] not	(35)
*sKvK o *sPvP	[sKvK] possible word in Afrikaans, [sPvP] not	(38)
IDENT[place] o *sTvT	[sTvT] possible word	(41)
{*sKvK, sPvP} o IDENT[place]	[sKvK], [sPvP] not possible words	(41)
*sTvT o cut-off	No variation	(42)

3.1.2 Motivation of analysis

The analysis presented above depends on two assumptions: (i) the existence of *sCvC-constraints; and (ii) that these constraints are ranked ||*sPvP o *sKvK o *sTvT||. In this section I discuss the evidence for these assumptions. The aim of this section is to show that these assumptions are well founded both cross-linguistically and theory internally. This section is structured as follows: In §3.1.2.1 I discuss the evidence in favor of the |coronal ™ velar ™ labial| place harmony scale in English. This serves as evidence for the ranking between the *sCvC-constraints. Section §3.1.2.2 motivates the existence of OCP-

constraints against multiple occurrences of [t], [k] or [p] within a single syllable. Section §3.1.2.3 then presents evidence that [s+stop]-structures are marked. Finally, in §3.1.2.4 I show that the *sCvC-constraints are just the local conjunction of the OCP-constraints with a constraint against [s+stop]-structures.

3.1.2.1 The place harmony scale in English

In the analysis above I assumed the ranking ||*sPvP o *sKvK o *sTvT||. Implicit in this ranking is the idea that labial place is more marked than velar place, which is again more marked than coronal place. This is expressed formally in the place harmony scale in (44).

(44) **Place harmony scale**

|coronal ™ velar ™ labial|

In this section I will provide evidence that (44) is the correct representation of the place harmony scale for English. I begin by presenting evidence that coronals are the most harmonic, establishing that |coronal ™ {velar, labial}| holds of English. I then discuss the evidence that velars are more harmonic than labials, establishing that |velar ™ labial|.

It is generally accepted that coronal place is universally less marked than velar and labial place (de Lacy, 2002, Jakobson, 1968, Lombardi, 2001, Paradis and Prunet, 1990, 1991, Prince, 1997, 1998). On these grounds it can then be expected that the ordering |coronal ™ {labial, velar}| should hold of English too. However, there is also more direct evidence from phonotactic restrictions in English that coronals are less marked than velars and labials.

There are certain sequences that are allowed with coronal consonants, but not with velar or labial consonants. Of course, the [sCvC]-forms that are the focus of the discussion represent one such example. But there are more. English allows word final clusters of up to four consonants. The first and second consonant can be of any place of articulation. However, the third and fourth position can only be filled by coronals (Fudge, 1969). This shows that the restrictions on the distribution of coronals are less strict than the restrictions on the distribution of velars and labials.

(45) **Word final consonants in English**

	Position 1	**Position 2**	**Position 3**	**Position 4**
Coronal	*cat*	*lint*	*pasts*	*sixths*
Velar	*kick*	*crank*	–	–
Labial	*dam*	*lamp*	–	–

In order to determine the harmonic ordering between velars and labials, we can also use phonotactic restrictions. There are certain sequences that are allowed with coronals and velars but not with labials. For instance, sequences of the form [sCvNC] are allowed with coronal and velars but not with labials – where both C's are voiceless homorganic stops and N is a nasal homorganic to the stops. This shows that labials are more restricted in their distribution than either velars or coronals.

(46) **Words of the form [sCvNC]**

Coronals	*stint*	*stunt*
Velars	*skink*	*skunk*
Labials	**spimp*	**spump*

The co-occurrence patterns of homorganic consonants in the English lexicon also suggests that labials are more marked than velars, which are again more marked than coronals. Using a dictionary with approximately 20,000 English words, Berkley (1994a, 1994b, 2000) calculated the number of words with two labials, two velars, and two coronals. She then calculated the number of such words that would have been expected had the consonants been allowed to combine freely. The observed/expected ratio for each of these three places of articulation can then be calculated. An observed/expected ratio of smaller than 1 means that fewer words of that structure occur than what are expected had the consonants been allowed to combine freely.[40] Berkley found that words with two labials are more under represented than words with two velars, and that words with two velars are again more under represented than words with two coronals. This suggests that labials are subject to stronger occurrence restrictions that velars, which again are subject to stronger restrictions than coronals. The table in (47) is based on data extracted from Berkley (2000:23-30), and is representative of the patterns that Berkley found. This table contains the statistics about words with two labials, velars or coronals separated by one or two segments.

(47) **Co-occurrence patterns of two labials, velars or coronals separated by one or two segments**

	Example	**Observed**	**Expected**	**Observed/Expected**
Labials	*pop, pulp*	118	207.9	0.57
Velars	*cock, crack*	81	113	0.71
Coronal	*ten, tact*	1148	1271.5	0.90

[40] In this she follows a method first used for Arabic by *inter alia* Pierrehumbert (1993) and Frisch *et al.* (2004).

These data show that all three kinds of words are under-represented (the observed/expected ratios for all three classes are smaller than 1). However, words with two labials are more under-represented than words with two velars, which is again more under-represented than words with two coronals.

In summary: (i) English restricts the occurrence of labials and velars more than the occurrence of coronals. This supports the harmonic ordering |coronal ™ {velar, labia}|. (ii) Similarly, English restricts the occurrence of labials more than velars, supporting this harmonic ordering |velar ™ labial|.

3.1.2.2 Constraints against the co-occurrence of homorganic voiceless stops

In the previous section we have seen evidence for the fact that English restricts the co-occurrence of homorganic voiceless stops in certain configurations. In order to capture this, I suggest the existence of OCP-constraints that penalize the co-occurrence of identical segments within a certain domain. In (48a) I give the general schema for these constraints, in (48b) the specific instantiations that apply to voiceless stops, and in (48c) the ranking that holds between these constraints in English.

(48) **OCP constraints**

a. **General** $*[\alpha \ldots \alpha]_\delta$: Do not allow two [α]'s in domain δ.

b. **Coronal** $*[t \ldots t]_\sigma$: Do not allow two [t]'s in one syllable.

Velar $*[k \ldots k]_\sigma$: Do not allow two [k]'s in one syllable.

Labial $*[p \ldots p]_\sigma$: Do not allow two [p]'s in one syllable.

c. **Ranking** $\|*[p \ldots p]_\sigma \circ *[k \ldots k]_\sigma \circ *[t \ldots t]_\sigma\|$

In the formulation of these constraints I assume that the domain over which they apply is the syllable. This domain can be defined differently. For instance, it might be that the domain should be morphologically defined as a word or a morpheme (Tessier, 2003, 2004). In the experiments that I will report in §3.2 below all tokens were of the form [sCvC]. These tokens were therefore all mono-syllabic, and there was no reason for subjects to assume that the tokens were not also mono-morphemic. For these specific tokens it would not matter whether the domain is defined as the syllable, the word or the morpheme. The constraint would apply to these forms under any of these definitions. For this reason I will not consider the implications of the different possible definitions of the domain here.[41]

In the OT literature OCP-constraints are often defined as the local self-conjunction of markedness constraints (Alderete, 1996, 1997, Itô and Mester, 1998, Smolensky, 1995). The constraints in (48b) would then be stated as the local self-conjunction of constraints against a singleton coronal, velar or labial stop. For instance, in stead of $*[p \ldots p]_{\sigma}$ we would have $*[p]_{\sigma}^{2}$ (do not violate *[p] twice in the same syllable).

I opt not to adopt this definition of OCP-constraints. Viewing an OCP-constraint $*[\alpha \ldots \alpha]_{\delta}$ as the local self-conjunction of the constraint $*[\alpha]$ presupposes the existence of the constraint against a single occurrence of $[\alpha]$. If we redefined the constraints in (48b) in terms of local self-conjunction, then we are presupposing of the existence of the constraints *[p], *[k] and *[t]. The markedness of [k] and [p] can be motivated by referring to the universal harmony scale |coronal ™ {velar, labial}| (see §3.1.2.1). Since

[41] See also the discussion on domains in §3.1.2.4 below.

velars and labials are marked relative to coronals, [k] and [p] are marked relative to [t]. However, coronal appears as the most harmonic member of this scale, and Gouskova (2003) has recently shown that markedness constraints against the most harmonic member of a harmony scale results in incorrect typological predictions. The existence of *[t] is therefore questionable.

The intuition behind the formulation of the constraints in (48) can be stated as follows: even if a single instance of some structure is not marked, multiple occurrences of that structure in some localized domain might still be marked. Even though [t] is not individually marked, two occurrences of [t] within a single syllable is marked. This can be expressed in terms of a harmony scale: $|\alpha_\delta$ ™ $(\alpha \ldots \alpha)_\delta|$. This scale should be read as follows: one occurrence of the structure α within domain δ is more harmonic than two occurrences of α within domain δ. In Gouskova's interpretation of such harmony scales there will not be a constraint against the least marked member of the scale (no constraint against α), but there will be a constraint against all other members of the scale (against two occurrences of α in domain δ).

The only difference between the three constraints in (48b) is the place of articulation to which each refers. I therefore assume that their ranking reflects the place harmony scale |coronal ™ velar ™ labial| motivated in the previous section (§3.1.2.1). The constraints are ranked so that the constraint that refers to most harmonic place of articulation (coronal) is ranked lowest.

English tolerates violation of all three of these constraints. There are many words with two [t]'s, [k]'s or [p]'s in the same syllable (*tot*, *cake*, *pop*). This implies that these OCP-constraints should rank below all faithfulness that could be violated in order to avoid violating the OCP-constraints.

3.1.2.3 On the markedness of [s+stop]-structures

In §3.1.2.4 I will show that each of the *sCvC-constraints is formed by locally conjoining the OCP-constraints from (48b) with a constraint against [s+stop]-structures. The existence of the OCP-constraints has been motivated in the previous section. However, I still need to show that there is a constraint against [s+stop]-structures. In (49) I state this constraint, and the rest of this section is then dedicated to motivating the existence of this constraint.

(49) ***[s+stop]**

Do not allow a tautosyllabic [s+stop]-sequence.

There is general acceptance in the literature that [s+stop]-structures are marked. However, there is no agreement on the reason for their markedness. The approaches in the literature can be classified into two broad groups: (i) *Cluster approach*. Under this approach these sequences are interpreted as consonant clusters. The markedness of the sequences (at least in word initial position) then results from the fact that they violate the sonority sequencing principle (Selkirk, 1982a). (ii) *Complex segment approach*. Under this approach an [s+stop]-structure is interpreted as a single complex segment rather than a sequence of two separate segments. The markedness of the structure then follows from the fact that it is complex segment (Sagey, 1990).

In order to explain the results of the experiments that I report in §3.2 the reason for the markedness of the [s+stop]-structures is not relevant. What is relevant is that these structures are marked, and that there is a constraint against them. I will therefore not choose between the two accounts for the markedness of these sequences. The rest of this section consists of discussion of a representative sample of the arguments in favor of both views on the markedness of [s+stop]-structures. The purpose of this discussion is to show that these structures are marked, whether they are viewed as consonant clusters or as complex segments.[42]

Cluster approach.[43] The first argument in favor of interpreting [s+stop]-structures as consonant clusters rests on English stress placement rules and comes from Hayes (1980, 1982, 1985). English regularly stresses the ante-penultimate syllable. However, when the penultimate syllable is heavy, stress is attracted from the ante-penult to the heavy penult. Hayes points out that words with an [s+stop]-structure preceding the final vowel are stressed as if the penultimate syllable is closed – i.e. stress falls on the penult. This can be explained if we assume that the [s] syllabifies into the coda of the penultimate syllable and the [stop] into the onset of the ultimate syllable. However, if a syllable boundary intervenes between the [s] and the [stop], then this structure should be interpreted as a sequence of two consonants. The examples in (50) illustrate this point.

---

[42] In fact, it is quite possible that both approaches are correct. It is possible that [s+stop]-structures are clusters in some languages and complex segments in others. It is even possible within one language that not all [s+stop]-structures have the same structural representation. See Fleischhacker (2001, 2002) for arguments to this effect.

[43] For examples of proponents of this approach, see *inter alia* Davis (1984), Hayes (1985:1480149), Hockett (1955:152-153), Gouskova (2001), Kahn (1980:42-43).

(50) **Stress placement in English**

Penult light = stressed ante-penult	**Penult heavy = stressed penult**
Cánada	*agénda*
sýllable	*patáto*
prínciple	*amálgam*

[s+stop] before final vowel acts like heavy penult words

seméster	**sémester*
Damáscus	**Dámascus*
Gilléspie	**Gíllespie*[44]

Closely related to this argument based on stress placement, is the argument based on the phenomenon of "pre-stress destressing" in English. This argument is also due to Hayes (1980, 1982, 1985). In pre-stress position English regularly de-stresses and consequently neutralizes vowels. However, the neutralization is blocked if the pre-stress vowel occurs in a closed syllable. Hayes points out that a pre-stress vowel followed by an [s+stop]-structure is also spared from neutralization. This can be explained if we assume that the [s] syllabifies into the coda of the pre-stress syllable, and the [stop] into the onset of the stressed syllable. Again, the intervening of a syllable boundary between the two parts of the structure is easier to explain if we assume that these structures are in fact consonant clusters. The examples in (51) illustrate this point.

---

[44] But see Davis (1982) and Lamontagne (1993) for counter examples such as *pédestal*, **pedéstal* and *áncestor*, **ancéstor*. According to Davis there are just about an equal number of words like *pédestal* and words like *seméster*.

(51) **Pre-stress destressing and neutralization in English**

Pre-stress open = neutralization	**Pre-stress closed = no neutralization**
s[ə]m*ester*	*d*[ɔ]*ctórial*
[ə]*mérica*	*c*[æ]*ntéén*
C[ə]*nnécticut*	*sh*[æ]*mpóó*

Pre-stress vowels followed by [s+stop]-sequences resist neutralization

pl[æ]*stícity*	**pl*[ə]*stícity*
m[æ]*scára*	**m*[ə]*scára*
m[æ]*stítis*	**m*[ə]*stítis*[45]

More evidence for the cluster approach comes from trends in syllabification. Treiman (1983), Treiman and Danis (1988) and Treiman *et al.* (1992) conducted a series of experiments in which they investigated how English speakers syllabify intervocalic consonant clusters. In general they found that, if two inter-vocalic consonants would form a licit onset, English speakers prefer to syllabify the consonants together into the onset of the second syllable – i.e. when presented with a token such [nəprim] speakers more often syllabified as [nə.prim] than as [nəp.rim]. However, they found the opposite for [s+stop]-structures. Speakers more often placed a syllable boundary between the [s] and the [stop] – i.e. when presented with a token such as [nəspim], speakers more often syllabified as [nes.pim] than as [ne.spim]. This suggests that the [s] and [stop] parts of these structures

[45] Again see Davis (1982) and Lamontagne (1993) for counter examples such as [ə]*stónish* and *n*[ə]*stálgia*.

are not very closely connected, and therefore than they are clusters rather than complex segments.

There is also evidence from languages other than English that [s+stop]-structures act like clusters rather than single segments. I will discuss an example from Italian here. This example is due to Kaye, Lowenstamm and Vergnaud (1990). Italian has a process that lengthens stressed vowels in open syllables. Stressed vowels in closed syllables, however, do not undergo this process. Stressed vowels followed by an [s+stop]-structure are also immune to the lengthening process. This can be understood if we assume that the [s] is syllabified into the coda of the stressed syllable and the [stop] into the onset of the following syllable. Again, the occurrence of a syllable boundary between the [s] and the [stop] can be explained more easily by assuming that these structures are consonant clusters rather than single complex segments.

(52) **Stressed syllable lengthening in Italian**

Open = lengthening		**Closed = no lengthening**	
f[á:]*to*	‘fate’	*f*[á]*tto*, **f*[á:]*tto*	‘fact’
m[é:]*ro*	‘pure’	*m*[á]*nto*, **m*[á:]*nto*	‘coat’

Pre-[s+stop] = no lengthening

p[á]*sta*, **p*[á:]*sta*	‘pasta’
p[é]*sta*, **p*[é:]*sta*	‘trail’

There is therefore ample evidence, both in English and cross-linguistically, that [s+stop]-structures are (sometimes) treated as consonant clusters. If we assume this structure for word initial [s+stop]-structures in English, then the markedness of these

structures can be explained easily. The "sonority sequencing principle" dictates that the sonority of consonants in the onset of a syllable cannot fall towards the nucleus (Selkirk, 1982a). The fricative [s] is higher is sonority than a stop (Parker, 2002). Consequently, an [s+stop] onset cluster violates the sonority sequencing principle.

Complex segment approach.[46] The first two arguments in favor of the complex segment approach involve syllable structure in English and come from Selkirk (Selkirk, 1982b). If we assume that [s+stop]-structures are clusters, then these clusters represent the only onsets in English that violate the sonority sequencing principle. All other onset clusters rise in sonority towards the nucleus, but [s+stop]-clusters fall in sonority. If [s+stop]-structures are interpreted rather as complex segments, then there are no exceptions to the sonority sequencing principle for onsets in English.

Closely related to this is the argument about the number of consonants that can occur in onset position in English. If we interpret [s+stop]-structures as complex segments then onsets in English are maximally two segments long. However, if [s+stop]-structures are interpreted as consonant clusters, then English does allow tri-consonantal onset clusters (cf. *street*, *splash*, *squeak*). All tri-consonantal onset clusters then have the structure [s + stop + glide/liquid]. Interpreting [s+stop]-structures as clusters therefore results in a complication to the syllable structure of English (tri-consonantal onsets are now allowed), and in an inexplicable idiosyncrasy of the tri-consonantal clusters (why [s + stop + glide/liquid] is the only observed structure). Both of these can be avoided by assuming that [s+stop]-structures are complex segments rather than consonant clusters.

---

[46] For examples of proponents of this approach, see *inter alia* Broselow (1991), Firth (1935, 1935-1937:543), Fudge (1969), Fujimura and Lovins (1978:112), Lamontagne (1993: Chapter 6), Selkirk (1982b), Spang-Hanssen (1959:158-161).

Also for the complex segment approach, evidence from languages other than English is available. I will discuss two examples here. First, about alliteration in Late-Middle and Early-Modern Irish verse and Old Germanic verse. These traditions made extensive use of alliteration as a poetic device. The rules determining what sounds can alliterate with what sounds were complicated. However, the rules in (53) capture the basic patterns (Kurylowicz, 1971, Meyer, 1909, Murphy, 1961).

(53) **Rules of alliteration in Irish and Germanic verse**

(i) Any vowel alliterated with any other vowel.

(ii) A consonant alliterated only with itself.

(iii) But [s+stop]-initial words could not alliterate with [s]-initial words. [st-] alliterated only with itself, and similarly for [sk-] and [sp-].

The peculiar behavior of [s+stop]-initial words can be explained more easily if it is assumed that the [s+stop]-structures formed single complex segments.

Another example of a language where [s+stop]-structures are treated as complex segments, is Egyptian Arabic. Egyptian Arabic allows no onset clusters. When Egyptian Arabic borrows words from English that start on a consonant clusters, vowels are inserted in order to break up the word initial consonant cluster. However, when English words that start on [s+stop] are borrowed into Egyptian Arabic, this does not happen. The examples in (54) are from Broselow (1991).

Unlike other consonant sequences, the sequence [s+stop] cannot be broken up by an epenthetic vowel. This is more easily understandable if we assume that these sequences are interpreted as single complex segments rather than consonant clusters.

(54) **English borrowings into Egyptian Arabic**

#CC → #CvC		**#s+stop → # ʔ i s+stop**	
plastic	[bilastik]	*ski*	[ʔiski]
Fred	[fired]	*study*	[ʔistadi]
sweater	[siwetar]	*spring*	[ʔispriŋ]
slide	[silaid]	*street*	[ʔistiri:t]

We therefore have evidence both from within English and cross-linguistically that [s+stop]-structures are (at least sometimes) treated as single complex segments. If we assume this interpretation for [s+stop]-structures, then they are marked simply by virtue of being complex segments. Complex segments have either multiple places of articulation or multiple manners of articulation (or both), and are as such marked relative to ordinary simplex segments. It is also true that complex segments are relative scarce in the segmental inventories of the world's languages (Ladefoged and Maddieson, 1996, Maddieson, 1984). This also serves as motivation for their markedness.

Whether we interpret word initial [s+stop]-structures in English as consonant clusters or as complex segments, it is clear that they are marked. If they are interpreted as consonant clusters, then they are marked by virtue of violating the sonority sequencing principle. If they are interpreted as complex segments, then they are marked *per se*. There must therefore exist some constraint that is violated by these structures. I use *[s+stop] from (49) to represent this constraint.

3.1.2.4 *sCvC-constraints as local conjunctions

In this section I will argue that each of the *sCvC-constraints is formed via the local conjunction (Smolensky, 1995) of OCP-constraints in (48b) with the constraint *[s+stop] from (49). Local conjunction is a well motivated algorithm for constructing complex constraints from simple constraints. The constraints that need to be conjoined to form the *sCvC-constraints are also well motivated (see discussion in the previous two sections). Both the method of constructing the *sCvC-constraints and the components that go into their construction are therefore well-motivated.

We have now motivated the existence of the OCP-constraints ||*[p … p]$_\sigma$ o *[k … k]$_\sigma$ o *[t … t]$_\sigma$|| and of the constraint *[s+stop]. English tolerates violation of each of these constraints individually, and even simultaneous violation of *[t … t]$_\sigma$ and *[s+stop]. What English does not tolerate is simultaneous violation of *[s+stop] and *[p … p]$_\sigma$ or *[k … k]$_\sigma$. A few examples are given in (55).

This is not an unknown phenomenon. There are countless examples of languages that tolerate violation of individual markedness constraints, but not the violation of certain combinations of markedness constraints in some local context. For instance, Itô and Mester (1997, 1998) analyze coda devoicing in German in this manner. German (and also Dutch, Afrikaans, Russian, etc.) tolerates both coda consonants and voiced obstruents separately, but not voiced obstruents in coda position. This generalization can be stated as follows in terms of markedness constraints: German tolerates violation of the constraint against having consonants in coda position (NoCoda) and of the constraint against voiced obstruents (*VoicedObs). What German does not tolerate, is

simultaneous violation of both of these constraints in one syllabic position – i.e. German does not tolerate violation of NoCoda and *VoicedObs in the same coda. This then explains why a word such as *Rades* [ra:.dəs] 'wheels' is well-formed in German, while a word such as *[ra:d] is not. The form [ra:.dəs] violates *VoicedObs because of the [d]. It also violates NoCoda because its final syllable is closed. However, these two constraints are not violated in the same coda. The form *[ra:d] also violates both *VoicedObs and NoCoda. This form, however, is ill-formed because these two constraints are both violated in the same coda position.

(55) **Violation of OCP-constraints and *[s+stop] in English**

a. Tolerated

Example	**Constraints violated**
sting	*[s+stop]
skate	*[s+stop]
speak	*[s+stop]
tot	*[t … t]$_\sigma$
cock	*[k … k]$_\sigma$
pop	*[p … p]$_\sigma$
state	*[s+stop] & *[t … t]$_\sigma$

b. Not tolerated

**skak*	*[s+stop] & *[k … k]$_\sigma$
**spap*	*[s+stop] & *[p … p]$_\sigma$

The restriction that English places on [sCvC]-words is therefore not unusual. Both Alderete (1996, 1997) and Itô and Mester (1998) suggested that restrictions like these can be analyzed via local conjunction of markedness constraints. The locally conjoined constraint is then violated only if all the individual constraints that are conjoined to form the conjoined constraint are violated simultaneously. The three *sCvC-constraints can now be shown to be formed from the local conjunction of an OCP-constraint with the constraint against [s+stop]-structures. These constraints are stated in these terms in (56).

(56) ***sCvC-constraints as local conjunction**

sTvT = [[s+stop] & *[t … t]$_\sigma$]$_\sigma$

sKvK = [[s+stop] & *[k … k]$_\sigma$]$_\sigma$

sPvP = [[s+stop] & *[p … p]$_\sigma$]$_\sigma$

Itô and Mester (1998) formulate a principle that they call "ranking preservation". This principle requires the following: Let LC_1 and LC_2 be two constraints formed via local conjunction, and let C_1 be one of the conjuncts of LC_1 and C_2 one of the conjuncts of LC_2. If $\|C_1 \circ C_2\|$, then $\|LC_1 \circ LC_2\|$. I have argued in (48c) above that the OCP-constraints are ranked as follows for English: $\|$*[p … p]$_\sigma$ o *[k … k]$_\sigma$ o *[t … t]$_\sigma\|$. *[p … p]$_\sigma$ is one of the conjuncts of *sPvP, *[k … k]$_\sigma$ of *sKvK, and *[t … t]$_\sigma$ of *sTvT. If we assume the principle of ranking preservation, it then follows that the three *sCvC constraints are indeed ranked as I assumed in §3.1.1 – i.e. $\|$*sPvP o *sKvK o *sTvT$\|$.

Locally conjoined constraints are *locally* conjoined – this means that they are evaluated in some local domain. It is not the case that the locally conjoined constraint is always violated when all of its conjuncts are violated, but only when all of its conjuncts

are violated within some local domain. In (56) I make the assumption that the relevant domain is the syllable. As with the OCP-constraints this domain could be stated differently (see the discussion in §3.1.2.2). The domain could also be defined morphologically as the word or the morpheme. In the experiments that I will discus below in §3.2 the tokens were of the form [sCvC]. All tokens were therefore monosyllabic, and there was no reason for subjects to assume that the tokens were poly-morphemic. Irrespective of whether we define the domain of the *sCvC-constraints as the syllable, the morpheme or the word, all of the tokens in the experiment would violate one of the *sCvC-constraints. For our purposes it is therefore not crucial which of these options we choose. I am making the arbitrary choice of assuming that the syllable counts as the domain.

I have shown above: (i) that the *sCvC-constraints are formed by the local conjunction of constraints that are all individually motivated both in English and cross-linguistically, and (ii) that the ranking ||*sPvP o *sKvK o *sTvT|| is based on the place harmony scale |coronal ™ velar ™ labial|. The basic OT analysis developed in §3.1.1 is therefore well-motivated. In the next section, I consider the predictions that follow from this analysis with regard relative well-formedness judgments and lexical decision reaction times.

3.1.3 The processing of non-words of the form [sCvC]

In §3.1.1 I developed the following mini-grammar for English: ||*sPvP o *sKvK o IDENT[place] o *sTvT o Cut-off||. Assuming that this analysis is correct, what predictions does it make about how English speakers will process non-words of the form [sCvC]? Recall the basic assumption about how grammar influences well-formedness judgments

and lexical decision reaction times (§1.1): (i) The more well-formed a token is according to the grammar of the language, the more well-formed it will be judged to be by language users. (ii) The more well-formed a non-word is according the grammar of the language, the more seriously language users will consider it as a possible word, and the longer they will take to reject it as a non-word.

In the experiments that I conducted, I included non-words of the form [sTvT], [sKvK] and [sPvP]. What underlying representation would the subjects have assumed for these tokens? Consider an [sPvP]-token as an example. If subjects assumed an underlying representation identical the token, then they will assume the mapping /sPvP/ → [sPvP]. This mapping will violate at least *sPvP. If they assumed any underlying representation other than /sPvP/, the map /input/ → [sPvP] will violate in addition to *sPvP also some faithfulness constraint. Say that the subjects assumed the underlying representation /sPvT/. The mapping /sPvT/ → [sPvP] will then violate both *sPvP and IDENT[place]. The assumption that the underlying representation is identical to the perceived token therefore results in the more harmonic mapping. For this reason, I will assume that the subjects assumed an /sPvP/-input for all [sPvP]-tokens. The same argument can be made for [sKvK]-tokens and [sTvT]-tokens. Under these assumptions, these three kinds of non-words can be compared as shown in (57) in the mini-grammar developed above. Since all three tokens are assumed to be faithful candidates, the faithfulness constraint IDENT[place] is not violated.

(57) **Comparing non-words**

	*sPvP	*sKvK	IDENT[place]	*sTvT
1 [sTvT]				*
2 [sKvK]		*		
3 [sPvP]	*			

Output of EVAL

Cut-off

sTvT

sKvK

sPvP

[sTvT]-forms violate the lowest ranked constraint and are therefore rated best by the grammar. [sPvP]-forms violate the highest ranked constraint and are therefore rated worst by the grammar. [sKvK]-forms occupy the intermediate position. From this we get the following predictions about well-formedness judgments: [sTvT]-forms will be rated as most well-formed, then [sKvK]-forms, and finally [sPvP]-forms. About lexical decision reaction times, the following predictions can be made: [sPvP]-forms will be rejected most quickly, [sKvK]-forms more slowly, and [sTvT]-forms will have the slowest rejection times. These predictions are represented graphically in (58).

(58) **Predictions with regard to well-formedness judgments and lexical decision**

Well-formedness judgments	**Rank-ordering imposed by EVAL**	**Lexical decision RT's**
Decreasing well-formedness judgments ↓	sTvT sKvK sPvP	↓ Decreasing reaction time in lexical decision

In the next section I discuss a series of well-formedness judgment and lexical decision experiments that I conducted to test these predictions. The results of these

experiments confirmed that the restriction on [sCvC]-words have an influence on how English speakers process non-words of this form. In particular, the results confirm the predictions shown in (58).

3.2 Word-likeness and lexical decision in English

Inspired by the experiments conducted by Berent *et al*. (see §2.2) for Hebrew, I conducted a similar set of experiments for English. These included both well-formedness judgment experiments and lexical decision experiments. In the experiments, subjects did indeed respond according to the predictions that follow from the analysis developed in §3.1 above. In the well-formedness judgment experiments, subjects rated the possible words ([sTvT]) better than both kinds of impossible words ([sKvK] and [sPvP]). They also distinguished between the two kinds of impossible words, rating [sKvK]-forms better than [sPvP]-forms. In lexical decision they distinguished between possible and impossible words, and between different kinds of impossible words. Non-words of the form [sTvT] were rejected slower than non-words of the form [sKvK] and [sPvP]. But the impossible non-words [sKvK] were also rejected slower than the impossible non-words [sPvP]. These experiments therefore confirmed the predictions of the analysis developed above in §3.1. But the experiments also show that language users make finer well-formedness distinctions than simply between grammatical/possible and ungrammatical/impossible. They also distinguish in the set of impossible words between the more and the less well-formed. Together with the results from Hebrew where we saw that language users distinguish between possible words in terms of well-formedness, these results show that grammar must be able to make multi-level well-formedness distinctions. In the rest of this section I first discuss the well-formedness judgment

experiments (§3.2.1), and then the lexical decision experiments (§3.2.2) that I conducted. Before getting into the details of the individual experiments, I will mention those aspects of the experimental design that were shared between the three experiments.

Recordings

The tokens for all experiments were recorded as read by a phonetically trained female speaker of standard American English. Recordings were made directly onto CD in a sound-proofed room. Each token was cut from the speech stream and stored electronically. The intensity of all tokens was equalized.

Token selection

An assumption that underlies all of the experiments that I discuss below is that grammar influences how language users perform well-formedness judgments and lexical decision tasks. This assumption requires that we control for the possible influence of lexical statistics that we know to influence both of these tasks (see the discussion in §1.2 above). The most rigorous control would be to select tokens such that lexical statistics and grammar conflict. Consider the comparison between [sTvT]-forms and [sKvK]-forms. According to the grammatical analysis developed above (§3.1), [sTvT] is less marked than [sKvK]. Grammar favors [sTvT], and we are expecting that [sTvT]-forms should be rated as more well-formed than [sKvK]-forms and that [sTvT]-forms should be detected as non-words more slowly than [sKvK]-forms. If we selected tokens such that the lexical statistics favor [sKvK]-forms over [sTvT]-forms, then lexical statistics would make the opposite predictions – [sKvK] should be rated better and detected more slowly as non-words. If in spite of such a conflict we found that subjects responded according to the

predictions of grammar, it could be interpreted as strong evidence that grammar rather than lexical statistics determined their responses.

A less rigorous control for the possible influence of lexical statistics would be to select tokens such that the lexical statistics between tokens types do not differ significantly. Consider again the comparison between [sTvT]-forms and [sKvK]-forms. If the lexical statistics of the [sTvT]-tokens and the [sKvK]-tokens do no differ from each other, then we cannot attribute to lexical statistics any difference in the way in which these tokens types are responded to. If we found that in such a situation subjects do rate [sTvT]-tokens better and detect them more slowly as non-words, we can attribute this to the influence of grammar.

In the selection of tokens for the experiments I first tried to use the more rigorous control. I resorted to the less rigorous control only when this was not possible.

Subjects

The same subjects participated in all three experiments. The subjects were 20 undergraduate students from the University of Massachusetts. All subjects were native speakers of American English, and none of them reported any speech or hearing disabilities. Subjects took part in the experiment for course credit in an introductory Linguistics class. Subjects were tested individually or in groups of up to four.

Order of experiments

There are three experiments and therefore six different orderings possible between the experiments. The order in which the experiments were presented was varied between

subjects, so that each of the six possible orderings was presented to roughly an equal number of subjects.

3.2.1 Well-formedness judgment experiments

Following Berent *et al.* I conducted two kinds of well-formedness judgment experiments, namely a gradient well-formedness judgment experiment and a comparative well-formedness judgment experiment. In the gradient well-formedness judgment experiment subjects are presented with individual tokens and rate each token for its word-likeness/well-formedness on a 5-point scale. In the comparative well-formedness judgment experiment subjects are presented with pairs of tokens and select the member of a pair that they deem most well-formed.

3.2.1.1 Gradient well-formedness judgment experiment

In this experiment I presented subjects with a list of tokens that contained non-words of the form [sTvT], [sKvK] and [sPvP]. Subjects were required to rate each token on a 5-point scale where a score of [5] corresponded to a form that they deemed to be very well-formed or very likely to be included in the lexicon of English. A score of [1] corresponded to a form that was not well-formed at all and that was very unlikely to ever be included in the lexicon of English. The three conditions in this experiment and the predictions in each condition are summarized in (59).

The first two conditions represent a comparison between possible words ([sTvT]) and impossible words ([sKvK] and [sPvP]), and the prediction is that the possible words will be rated better than the impossible words. The more interesting comparison is the third one, where the two kinds of impossible words are compared. The prediction there is

that, although neither [sKvK] nor [sPvP] is a possible word of English, [sKvK] will be rated better than [sPvP]. Below I first discuss the experimental design (§3.2.1.1.1) and then the results of the experiment (§3.2.1.1.2).

(59) **Gradient well-formedness: Conditions and predictions**

Condition	**Prediction**
T~K	[sTvT] rated better than [sKvK]
T~P	[sTvT] rated better than [sPvP]
K~P	[sKvK] rated better than [sPvP]

3.2.1.1.1 Experimental design

Token selection

In this experiment we are looking for the influence of grammar on well-formedness judgments. It is therefore important to control for other factors that are known to influence well-formedness judgments. In particular, we have to control for the influence of lexical statistics (see discussion in §1.2.1). Non-words that inhabit denser lexical neighborhoods are usually rated as more well-formed than non-words that inhabit sparser lexical neighborhoods. Similarly, non-words with higher phoneme transitional probabilities and/or have higher phoneme transitional probabilities are rated as more well-formed. In the token selection I controlled for the influence of these two kinds of lexical statistics.

Consider first the T~K-condition. In this condition I selected five non-words of the form [sTvT] and five non-words of the form [sKvK]. For each non-word I calculated

a lexical neighborhood density[47] and a cumulative bi-phone transitional probability.[48, 49] These tokens were selected such that the mean lexical neighborhood density of the five [sTvT]-tokens did not differ significantly from the mean lexical neighborhood density of the five [sKvK]-tokens. Similarly, the mean cumulative bi-phone probability of the two kinds of tokens did not differ significantly. Now consider the T~P-condition. In this condition I selected five non-word tokens of the form [sTvT] and five non-word tokens of the form [sPvP]. These tokens were also selected such that lexical statistics did not favor any of the token types. Finally, consider the K~P condition. For this condition five non-words of the form [sKvK] and five non-words of the form [sPvP]-were selected. The lexical neighborhood density of the [sKvK]-tokens and [sPvP]-tokens do not differ significantly. However, the [sPvP]-tokens had significantly higher cumulative bi-phone probabilities than the [sKvK]-tokens. In this condition, cumulative bi-phone probabilities therefore favor the [sPvP]-tokens. The table in (60) shows the mean lexical statistics of

---

[47] The lexical statistics were all calculated from the CELEX database (Baayen *et al*., 1995). This database was "Americanized" before the calculations were done. In the "Americanization" the phonetic transcription of the forms in the database were changed so that they reflect standard American rather than British pronunciation. I am indebted to John Kingston for this.

Lexical neighborhood density was calculated according to the method used by *inter alia* Vitevitch and Luce (1998, 1999) and Newman *et al*. (1997). The neighbors of a token are defined as any word that can be formed from the token by substitution, addition or deletion of one phoneme from the token.

Lexical neighborhood density is calculated as follows: (i) Find all the neighbors for a token. (ii) Sum the log frequencies of all the neighbors. The lexical neighborhood density therefore takes into account both the number of neighbors and their frequencies.

[48] Transitional probabilities were also calculated form the CELEX database (Baayen *et al*., 1995), following the method used by Vitevitch and Luce (1998, 1999) and Newman *et al*. (1997). Consider the token [sKɔK] as an example. For the sequence [sK] we can calculate the probability of an [s] being followed by [K], and the probability of a [K] being preceded by an [s]. To calculate the probability of [s] being followed by [K]: (i) add up the log frequencies of all tokens that contain an [s]; (ii) add up the log frequencies of all tokens that contain the sequence [sK]; (iii) divide the log frequency of [sK] by the log frequency of [s]. The probability of a [K] being preceded by a [s] can be calculated in a similar manner.

[49] The cumulative bi-phone probability is the product of the individual bi-phone probabilities for a token. This is intended as a measure of the overall probability of the token.

the token types in each of the conditions. (A list of all the tokens and the lexical statistics of each token can be found in the Appendix.)

(60) **Lexical statistics for gradient well-formedness judgment experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	27.43	23.31	$t(8) = 0.52$, two-tailed $p = 0.62$
CBP	8.86×10^{-9}	6.99×10^{-9}	$t(8) = 0.30$, two-tailed $p = 0.77$

b. T~P-condition

	[sTvT]	[sPvP]	
LND	27.43	22.77	$t(8) = 0.67$, two-tailed $p = 0.52$
CBP	8.86×10^{-9}	3.44×10^{-9}	$t(8) = 1.09$, two-tailed $p = 0.31$

c. K~P-condition

	[sKvK]	[sPvP]	
LND	17.47	22.77	$t(8) = 1.18$, two-tailed $p = 0.27$
CBP	6.46×10^{-10}	3.44×10^{-9}	$t(8) = 3.76$, one-tailed $p = 0.002$[50]

Procedure

The list of tokens that were presented to subjects was constructed as follows: There was a total of 24 test-tokens.[51] Each of these 24 test-tokens was included twice in the list. To

[50] This *p*-value is based on a one-tailed *t*-test, assuming that the cumulative bi-phone probability of the [sPvP]-tokens is higher than that of the [sKvK]-token.

this 77 fillers[52] were added so that the final list contained 125 tokens. The list was presented auditorily to the subjects twice so that each test-token was presented four times. There was a break of about five minutes between the two presentations of the list. The list was differently randomized on each presentation. After hearing a token, subjects indicated their rating of the token on an answer sheet by circling a number from [1] to [5]. A score of [1] corresponded to a token that was judged as not very well-formed/very unlikely to ever be included in the lexicon of English. A score of [5], on the other hand, corresponded to a token that was judged to be very well-formed/very likely to be included in the lexicon of English. After a lapse of 5 seconds, the next token was presented. Before the list was presented the first time, 10 filler tokens were presented as practice trials.

3.2.1.1.2 Results

The results were submitted to a repeated measure ANOVA, with mean rating as dependent variables, and markedness (marked~unmarked)[53] and condition (T~K, T~P, K~P) and as independent variables. A main effect of markedness was found both by subjects ($F(1, 19) = 25.01$, $p < 0.000$) and by items ($F(1, 4) = 136.69$, $p < 0.000$). There

---

[51] This number is smaller than the expected number of 30 (3 conditions × 2 token types per condition × 5 tokens per token type), because The reason for this is that the same token is sometimes used in two different conditions.

[52] The fillers were selected such that approximately an equal number of all tokens were possible words and impossible words. Fillers that represented impossible words violated a constraint on the consonants that co-occur in the onset and syllable of a single syllable (Fudge, 1969). They were therefore ill-formed for reasons similar to ill-formedness of the [sKvK] and [sPvP]-tokens.

[53] In each condition, the tokens that are more marked according to the grammar were coded as “marked” and the tokens that are less marked according the grammar were coded as “unmarked”. In the T~K-condition, for instance, [sTvT]-tokens were coded as “unmarked” and [sKvK]-tokens as marked.

was also a significant interaction between markedness and condition both by subject ($F(2, 18) = 19.17$, $p < 0.000$) and by items ($F(2, 3) = 51.69$, $p = 0.005$).

The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. (One-tailed tests are called for because I am testing an *a priori* hypothesis about the relationship between well-formedness ratings in each condition.) In the K~P-condition no significant difference was found between the marked [sPvP]-tokens and unmarked [sKvK]-tokens, either by subjects ($t(19) = 0.88$, $p = 0.19$) or by items ($t(8) = 1.29$, $p = 0.12$). However, in the T~K-condition, the unmarked [sTvT]-tokens were rated better than the marked [sKvK]-tokens both by subjects ($t(19) = 5.81$, $p < 0.000$) and by items ($t(8) = 13.61$, $p < 0.000$). Similarly, in the T~P-condition, the unmarked [sTvT]-tokens received higher ratings than the marked [sPvP]-tokens, both by subjects ($t(19) = 5.39$, $p < 0.000$) and by items ($t(8) = 13.40$, $p < 0.000$).

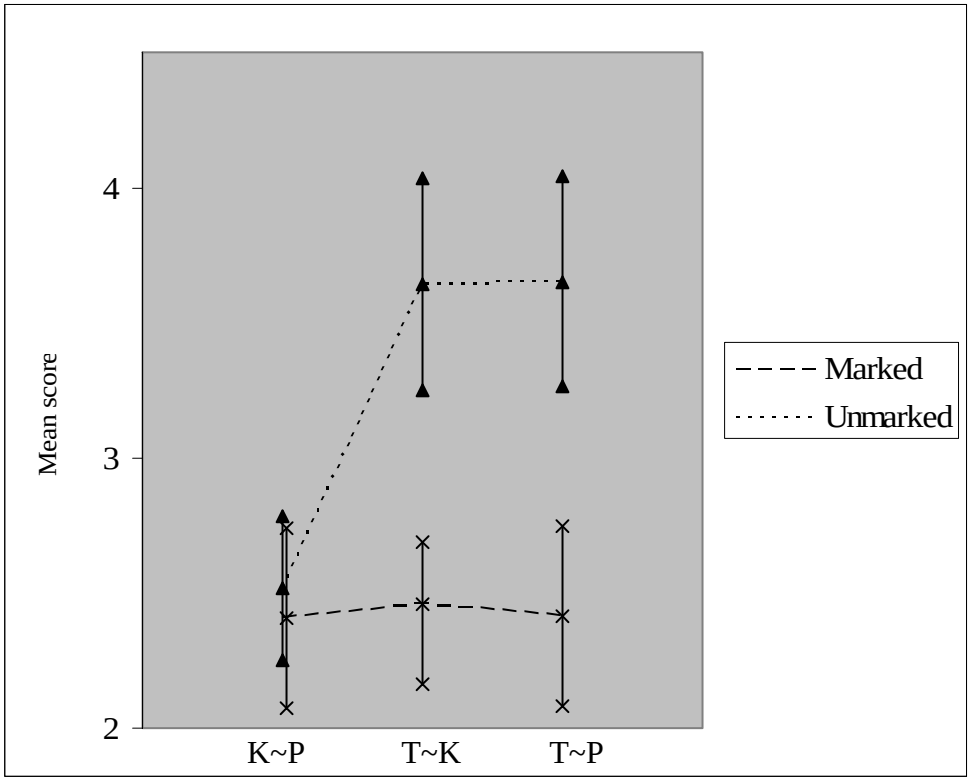
The figures in (61) represent the results graphically, and the results are also summarized in the tables in (62).

Discussion

In this experiment we have very strong evidence that non-words that are less marked according to the grammar are rated as more well-formed than non-words that are more marked. This is confirmed by the highly significant results on the main effect of markedness in the ANOVA's. More specifically, we also have unequivocal evidence that non-words of the form [sTvT] are rated as more well-formed than non-words of the form [sPvP] and [sKvK]. However, we do not have strong evidence for a preference for [sKvK]-tokens over [sPvP]-tokens. [sKvK]-tokens were rated better than [sPvP]-tokens on average, but the difference in ratings for these two kinds of tokens was small and did

not reach significance. Although there is not conclusive evidence that [sKvK] is rated as more well-formed than [sPvP], there is also no evidence to the contrary – i.e. [sPvP]-tokens were not rated better than [sKvK]-tokens. This is particularly relevant since the [sPvP]-tokens were favored over the [sKvK]-tokes in terms of the cumulative bi-phone probabilities (see table (60c) above). Had lexical statistics determined the response pattern, then [sPvP]-tokens should have been rated better. We also have no conclusive evidence that lexical statistics determined the response patterns in the [sKvK]~[sPvP]-condition.

(61) **Gradient well-formedness judgments: Mean scores by subject (with 95%-confidence intervals)**



(62) **a. Overall results (ANOVA's)**

	By subject		By item	
Main effect of markedness	$F(1,19) = 25.01$	$p < 0.000$	$F(1,4) = 136.69$	$p < 0.000$
Interaction with condition	$F(2,18) = 19.17$	$p < 0.000$	$F(2,3) = 51.69$	$p = 0.005$

b. Contrasts per condition

	K~P		T~K		T~P	
	[sKvK]	**[sPvP]**	**[sTvT]**	**[sKvK]**	**[sTvT]**	**[sPvP]**
Mean	2.52	2.41	3.64	2.43	3.65	2.41
By subject[54]	$t(19) = 0.88,\ p = 0.19$		$t(19) = 5.81,\ p < 0.000$		$t(19) = 5.39,\ p < 0.000$	
By item[55]	$t(8) = 1.29,\ p = 0.12$		$t(8) = 13.61,\ p < 0.000$		$t(8) = 13.40,\ p < 0.000$	

3.2.1.2 Comparative well-formedness judgment experiment

In this experiment I presented subjects with non-word token pairs. Each pair had one of the following forms: [sTvT]~[sKvK], [sTvT]~[sPvP], or [sKvK]~[sPvP]. The task of subjects was to select from each pair the member that they considered to be most well-formed/most likely to be included in the lexicon of English. The three conditions in this experiment and the predictions in each condition are summarized in (63).

As in the previous experiment, the first two conditions represent a comparison between possible words ([sTvT]) and impossible words ([sKvK] and [sPvP]), and the prediction is that the possible words will be preferred over the impossible words. The third condition is a comparison between two kinds of impossible words. The prediction there is that, although neither [sKvK] nor [sPvP] is a possible word of English, [sKvK]

---

54 *t*-test is paired sample of two means, *p*-values are one-tailed.

55 *t*-test is (non-paired) two sample of means, *p*-values are one-tailed.

will be preferred over [sPvP]. Below I first discuss the experimental design (§3.2.1.2.1) and then the results of the experiment (§3.2.1.2.2).

(63) **Comparative well-formedness: Conditions and predictions**

Condition	**Prediction**
T~K	[sTvT] preferred over [sKvK]
T~P	[sTvT] preferred over [sPvP]
K~P	[sKvK] preferred over [sPvP]

3.2.1.2.1 Experimental design

Token selection

Consider the T~K-condition as an example. I selected 15 non-word pairs of the form [sTvT]~[sKvK]. The members of a pair were selected so that the cumulative bi-phone probability of the more marked [sKvK]-member was higher than that of the less marked [sTvT]-member. For most of the pairs, the lexical neighborhood density of the more marked [sKvK]-token was also higher than that of the less marked [sTvT]-token.[56] The mean lexical neighborhood density and cumulative bi-phone probability of the more marked [sKvK]-tokens were higher than that of the less marked [sTvT]-tokens. The tokens in the other two conditions were selected in a similar manner. The mean lexical statistics of the tokens in each of the three conditions are given in (64). (The actual tokens used and their lexical statistics are given in the Appendix.)

[56] It was not possible to select 15 pairs in which both the lexical neighborhood density and the cumulative bi-phone probability of the more marked [sKvK]-token were higher than that of the less marked [sTvT]-token. Vitevitch and Luce (1998, 1999) have shown that transitional probabilities are more important than lexical neighborhood density in tasks that do not require lexical access. Since well-formedness judgments do not require lexical access, I decided that it was more important to control for transitional probability than for lexical neighborhood density.

This experimental design presents a rigorous test of the hypothesis that grammar in addition to lexical statistics influences well-formedness ratings. In all three conditions, lexical statistics favor the more marked tokens while grammar favors the less marked tokens. For instance, the lexical neighborhood density and cumulative biphone probability of the [sPvP]-tokens were higher than that of the [sTvT]-tokens. Based on this we would expect that [sPvP]-tokens will be rated better. In terms of markedness, the [sTvT]-forms were more well-formed. Based on this we would expect that subjects would rate the [sTvT]-tokens better. Lexical statistics and grammar conflict directly. If the responses of the subjects reflect the influence of grammar rather than lexical statistics, it would count as strong evidence that grammar does influence well-formedness ratings.

(64) **Lexical statistics for comparative well-formedness judgment experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	16.55	24.21	$t(14) = 3.32$, one-tailed $p < .003$
CBP	1.62×10^{-9}	8.15×10^{-9}	$t(14) = 3.43$, one-tailed $p < .003$

b. T~P-condition

	[sTvT]	[sPvP]	
LND	14.87	21.45	$t(14) = 4.30$, one-tailed $p < .001$
CBP	6.98×10^{-10}	3.17×10^{-9}	$t(14) = 6.21$, one-tailed $p < .000$

((64) continued)

c. K~P-condition

	[sKvK]	[sPvP]	
LND	10.30	21.26	$t(14) = 5.26$, one-tailed $p < .000$
CBP	2.58×10^{-10}	3.18×10^{-9}	$t(14) = 8.88$, one-tailed $p < .000$

Procedure

There were 15 token-pairs in each of the three conditions, resulting in 45 test-pairs. I added 45 filler pairs to this. Only non-words were used in the filler pairs.[57] This resulted in a total of 90 token-pairs. Two lists were created from these 90 token-pairs. Each list contained all 90 token-pairs. In List 1, eight out of the fifteen pairs of the T~K-condition had the [sTvT]-token first and the [sKvK]-token second. In the other seven token-pairs for this condition, the [sKvK]-token was used first. The same was true for the T~P-pairs and K~P-pairs. In List 2 the order between the members in a token pair was reversed – i.e. if two tokens occurred in the order [Token 1]~[Token 2] in List 1, then they occurred in the order [Token 2]~[Token 1] in List 2. Both lists were presented auditorily to subjects. About 5 minutes elapsed between the presentation of the lists. On each presentation of a list, it was differently randomized. Before the list was presented the first time, 10 filler token-pairs were presented as practice trials.

Subjects were instructed that they would hear a pair of non-words, and that their task would be to select the member of each pair that they thought could most likely be

[57] The fillers in this experiment were selected using the same criteria as in the previous experiment. See footnote 54.

included in the lexicon of English in the future. Subjects indicated their response by pushing one of two buttons. The left hand button was pushed if the first token in a pair was preferred, and the right hand button if the second token was preferred. Responses were recorded electronically. The next token pair was presented 500 ms after all subjects had responded, or after a time of 5 seconds has elapsed. If a subject did not respond within this time frame, a non-response was recorded.

3.2.1.2.2 Results

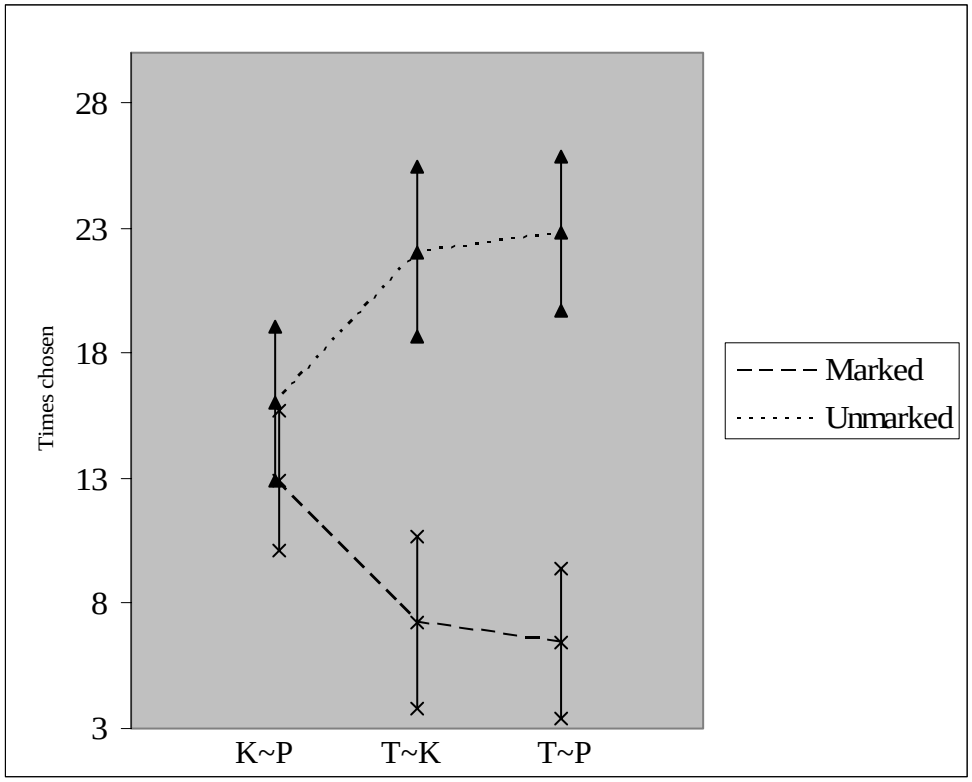
There were 15 token pairs per condition, each of which was presented twice. There were therefore 30 responses per subject per condition. On every response a subject could either select the more marked or the less marked member. These results were subjected to a 2 × 3 ANOVA with markedness (marked~unmarked) and condition (K~P, T~K, T~P) as independent variables. A main effect of markedness was found both by subjects ($F(1,19) = 23.28$, $p < 0.000$) and by items ($F(1,14) = 188.43$, $p < 0.000$). There was also a significant interaction between markedness and condition both by subjects ($F(2,18) = 10.37$, $p = 0.001$) and by items ($F(2,13) = 23.91$, $p < 0.000$).

The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. In the K~P-condition there was an advantage for the less marked [sKvK]-tokens over the more marked [sPvP]-tokens. This difference was significant by items ($t(14) = 1.92$, $p = 0.037$), but not by subjects ($t(19) = 1.12$, $p = 0.14$). In the T~K-condition, the unmarked [sTvT]-tokens were preferred over the marked [sKvK]-tokens both by subjects ($t(19) = 4.54$, $p < 0.000$) and by items ($t(14) = 15.58$, $p < 0.000$). Similarly, in the T~P-condition, the unmarked [sTvT]-tokens were preferred

over the marked [sPvP]-tokens, both by subjects ($t(19) = 5.73$, $p < 0.000$) and by items ($t(14) = 13.09$, $p < 0.000$).

The figures in (65) represent the results graphically, and the results are also summarized in the tables in (66).

(65) **Comparative well-formedness judgments: Mean number of choices for the marked and unmarked member in each condition by subject (with the 95%-confidence intervals)**



(66) **a. Overall results (ANOVA's)**

	By subject		By item	
Main effect of markedness	$F(1,19) = 23.28$	$p < 0.000$	$F(1,14) = 188.43$	$p < 0.000$
Interaction with condition	$F(2,18) = 10.37$	$p = 0.001$	$F(2,13) = 23.91$	$p < 0.000$

((66) continued)

b. Contrasts per condition

	K~P		T~K		T~P	
	[sKvK]	**[sPvP]**	**[sTvT]**	**[sKvK]**	**[sTvT]**	**[sPvP]**
By subject[58]	16.0	12.9	22.1	7.3	22.8	6.4
	$t(19) = 1.12$, $p = 0.14$		$t(19) = 4.54$, $p < 0.000$		$t(19) = 5.73$, $p < 0.000$	
By item[59]	21.3	17.2	29.4	9.7	30.4	8.5
	$t(14) = 1.92$, $p = 0.037$		$t(14) = 15.58$, $p < 0.000$		$t(14) = 13.09$, $p < 0.000$	

Discussion

Like the gradient well-formedness experiment, this experiment provides very clear evidence for the effect of markedness in general. This is confirmed by the highly significant results on the main effect of markedness in the ANOVA's. We also have very strong evidence that non-words of the form [sTvT] are preferred over non-words of the form [sKvK] and [sPvP] – this preference was found to be highly significant both by subjects and by items. In this experiment there is also evidence that non-words of the form [sKvK] are preferred over non-words of the form [sPvP]. In the K~P-condition there is an absolute preference for the [sKvK]-tokens both in terms of subjects and in terms of tokens. This preference is significant by items.

In all three of the conditions, the more marked tokens were favored by the lexical statistics. (See (64) above.) Had lexical statistics been responsible for these response patterns, we would have expected a preference for the more marked member. The fact

[58] *t*-test is paired sample of two means, *p*-values are one-tailed.

[59] *t*-test is paired sample of two means, *p*-values are one-tailed.

that the opposite preference was found in all three conditions is therefore very strong evidence that grammar influences well-formedness judgments. These results show in particular that grammar takes precedence when grammar and lexical statistics conflict.

Based on the grammatical analysis developed above, I hypothesized the following well-formedness relationship between [sCvC]-forms: |sTvT ™ sKvK ™ sPvP|. In both well-formedness judgment experiments we have clear evidence that non-words of the form [sTvT] are judged to be more well-formed than non-words of the form [sPvP] or [sKvK]. I therefore interpret these two experiments as strongly confirming the sub-hypothesis: |sTvT ™ {sKvK, sPvP}|. For the K~P-condition we found an absolute advantage for the less marked [sKvK]-tokens over the more marked [sPvP]-tokens in both experiments. In the gradient well-formedness experiment this preference did not reach significance. However, in the comparative well-formedness judgment experiment, it was found to be significant by items. This gives evidence for the sub-hypothesis |sKvK ™ sPvP|. In general I interpret the results of these experiments as confirmation of the hypothesized well-formedness relation between [sCvC]-forms. The lexical decision experiment discussed below provides additional confirmation of this hypothesis.

3.2.2 Lexical decision experiment

In this experiment I presented subjects auditorily with a list of words and non-words. The task of subjects was to discriminate between words and non-words. The basic hypothesis is that listeners use, among other things, the information provided by grammar when they make lexical decisions. The less well-formed a non-word token is, the less seriously a listener will consider it as a possible word, and the quicker the token will be rejected. The non-words that the listeners were presented with included tokens of the form [sTvT],

[sKvK] and [sPvP]. The three experimental conditions and the predictions for the conditions are summarized in (67) below.

(67) **Lexical decision: Conditions and predictions**

Condition	**Prediction**
T~K	RT([sTvT]) > RT([sKvK])
T~P	RT([sTvT]) > RT([sPvP])
K~P	RT([sKvK]) > RT([sPvP])

As in the previous experiments, the first two conditions represent a comparison between possible ([sTvT]) and impossible words ([sKvK] and [sPvP]), and the prediction is that the possible words will be rejected more slowly than impossible words. The third condition is a comparison between two kinds of impossible words. The prediction is that, although neither [sKvK]- nor [sPvP]-forms are possible words of English, [sKvK]-forms will be rejected more slowly than [sPvP]-forms. This was confirmed by the results of the lexical decision experiment. In the rest of this section I will first discuss the design (§3.2.2.1), and then the results (§3.2.2.2) of the lexical decision experiment.

3.2.2.1 Experimental design

Token selection

Consider the T~K-condition as an example. I selected 5 non-words of the form [sTvT] and 5 non-words of the form [sKvK]. The tokens were selected such that the mean lexical neighborhood density and the mean cumulative bi-phone probability of the [sTvT]-tokens and the [sKvK]-tokens did not differ significantly. Lexical statistics therefore did not favor any of the two token types. The token selection in the other two conditions was

done in a similar way. The mean lexical statistics of the tokens in each of the three conditions are given in (68). (The actual tokens used and their lexical statistics are given in the Appendix.)

(68) **Lexical statistics for comparative well-formedness judgment experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	31.58	27.24	$t(8) = 0.78$, two-tailed $p > .45$.
CBP	4.85×10^{-8}	6.18×10^{-9}	$t(8) = 1.81$, two-tailed $p > .11$

b. T~P-condition

	[sTvT]	[sPvP]	
LND	31.58	26.56	$t(8) = 1.00$, two-tailed $p > .34$
CBP	4.85×10^{-8}	2.54×10^{-9}	$t(8) = 1.99$, two-tailed $p > .08$

c. K~P-condition

	[sKvK]	[sPvP]	
LND	14.95	14.26	$t(8) = 0.10$, two-tailed $p > .92$
CBP	8.08×10^{-10}	8.50×10^{-10}	$t(8) = 0.06$, two-tailed $p > .95$

Procedure

The list of tokens that were presented to subjects was constructed as follows: There were a total of 27 test-tokens.[60] Each of these 27 test-tokens was included once in the list. To

[60] This number is smaller than the expected number of 30 (3 conditions × 2 token types per condition × 5 tokens per token type), because the same token is sometimes used in two different conditions.

this 76 fillers[61] were added so that the final list contained 103 tokens. These tokens were presented auditorily to subjects. On each presentation the list was differently randomized. The list was presented twice to subjects, with a break of about five minutes between presentations. Subjects responded by pressing one of two buttons on a response box. One button was marked as “Yes”, and was used to indicate that the token was a word of English. The other button was marked as “No”, and was used to indicate that the token was not a word of English. The order between the buttons was varied so that half of the subjects responded “Yes” with the right hands, and half responded “No” with their right hands. Subjects were instructed to respond as quickly as possible, but to listen to the whole token before responding. The next token was presented after all subjects have responded or after 2 seconds have elapsed. Both responses and response times were recorded. Before the list was presented the first time, 10 filler tokens were presented as practice trials.

3.2.2.2 Results

The response times were recorded starting at the onset of a stimulus. Before analyzing the data I subtracted the duration of every stimulus from the recorded response time. The resulting measure represents how long after (or before) the end of the stimulus a subject recorded a response. In the rest of the discussion I will refer to this measure (recorded response time minus token duration) as “response time”. The response times for each subject were normalized, and responses that were more than 2 standard deviations away from the mean for a subject were excluded from the analysis. Only correct non-word

---

[61] The fillers in this experiment were selected using the same criteria as in the previous experiments. See footnote 54.

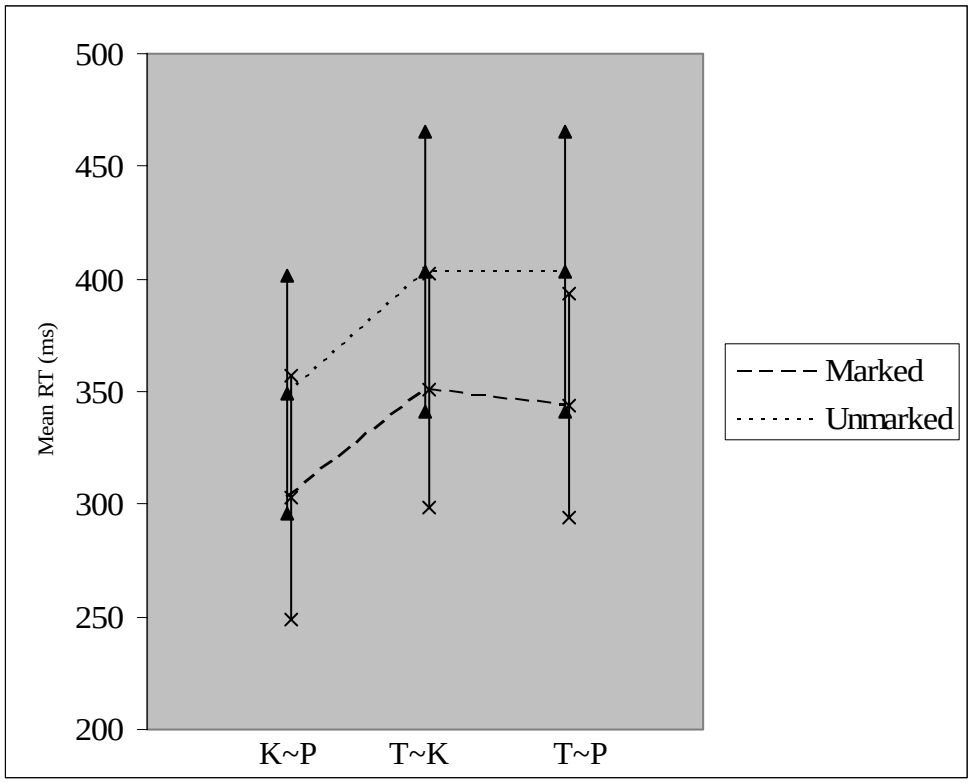
responses were included in the analysis. Exclusion of outliers and incorrect responses resulted in exclusion of only 8% of the total responses.

The response time data were subjected to a 2 × 3 ANOVA with markedness (marked~unmarked) and condition (K~P, T~K, T~P) as independent variables. A main effect of markedness was found by subjects ($F(1, 19) = 21.68$, $p = 0.001$), but not by items ($F(1, 4) = 2.584$, $p = 0.18$). There was no interaction between markedness and condition by subjects ($F(2, 18) = 0.58$, $p = 0.57$) or by items ($F(2, 13) = 0.08$, $p = 0.92$).

The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. In the K~P condition the more marked [sPvP]-tokens had shorter reaction times than the less marked [sKvK]-tokens. This difference was significant both by subjects ($t(19) = 3.79$, $p < 0.000$) and by items ($t(8) = 2.15$, $p = 0.03$). In the T~K-condition the more marked [sKvK]-tokens had shorter reaction times than the less marked [sTvT] tokens. This difference was significant by subjects ($t(19) = 3.40$, $p < 0.002$) but not by items ($t(8) = 1.63$, $p = 0.07$). In the T~P-condition the more marked [sPvP]-tokens were also rejected more quickly than the less marked [sTvT]-tokens. This difference was significant by subjects ($t(19) = 4.20$, $p < 0.001$) but not by items ($t(8) = 1.55$, $p = 0.08$).

The figures in (69) represent the results graphically, and the results are also summarized in the tables in (70).

(69) **Lexical decision: Mean RT's in ms by subject (with 95% confidence intervals)**



(70) **a. Overall results (ANOVA's)**

	By subject		By item	
Main effect of markedness	$F(1,19) = 21.68$	$p = 0.001$	$F(1,4) = 2.584$	$p = 0.18$
Interaction with condition	$F(2,18) = 0.58$	$p = 0.572$	$F(2,3) = 0.08$	$p = 0.92$

b. Contrasts per condition

	K~P		T~K		T~P	
	[sKvK]	**[sPvP]**	**[sTvT]**	**[sKvK]**	**[sTvT]**	**[sPvP]**
By subject[62]	348.98	303.33	403.53	350.45	403.53	344.01
	$t(19) = 3.78$, $p < 0.001$		$t(19) = 3.39$, $p < 0.002$		$t(19) = 4.20$, $p < 0.001$	
By item[63]	345.27	303.86	404.52	351.55	404.52	348.18
	$t(8) = 2.15$, $p = 0.03$		$t(8) = 1.62$, $p = 0.07$		$t(8) = 1.55$, $p = 0.080$	

[62] *t*-test is paired sample of two means, *p*-values are one-tailed.

[63] *t*-test is (non-paired) two sample of means, *p*-values are one-tailed..

Discussion

This experiment provides clear evidence that non-words that are more well-formed according to grammar are rejected more slowly than non-words that are less well-formed. This is confirmed by the significant main effect in the subjects ANOVA. It is also confirmed by the *t*-tests for each of the three individual conditions. In this experiment the lexical statistics did not differ significantly between the token types in each condition, so that it is not expected that the lexical statistics should significantly influence the reaction times. I therefore interpret the differences in reaction times that were found as evidence that grammar influences reaction times. Based on the grammatical analysis that I developed in §3.1 above, I hypothesized that the reaction time for the three kinds of non-words would be related as follows: RT([sTvT]) > RT([sKvK]) > RT([sPvP]). This hypothesis was confirmed by the results of the lexical decision experiment. Under the assumption that a non-word is rejected more quickly the less well-formed it is, this could be interpreted as evidence for the following well-formedness relation between the three kinds of [sCvC]-tokens: |sTvT ™ sKvK ™ sPvP|. The results of this experiment serve as strong evidence in favor of the grammatical analysis of these forms developed in §3.1 above.

4. Considering alternatives

In this section I consider alternative accounts for the response patterns observed in the well-formedness judgment and lexical decision experiments discussed above. Two kinds of alternatives are considered. First, I discuss the claim that the response patterns result from the influence of lexical statistics rather than from grammar. Secondly, I consider

several alternative grammatical accounts for the response patterns. I will show in particular that a rank-ordering model of EVAL is more suited to account for the response patterns than a classic OT grammar (Prince and Smolensky, 1993), a stochastic OT grammar (Boersma, 1998, Boersma and Hayes, 2001), or a grammar with variable constraint rankings (Anttila, 1997, Anttila and Cho, 1998, Reynolds, 1994, Reynolds and Sheffer, 1994).

4.1 Grammar or lexical statistics?

One of the central assumptions behind the interpretation of the experimental results in §2 and §3 above is that the response patterns observed in these experiments stem from the influence of grammar on phonological processing. However, it is firmly established that lexical statistics influence phonological processing (see the discussion in §1.2 above). Therefore, it is necessary to consider an alternative account for the response patterns – an account that attributes these patterns not to grammar but to lexical statistics. Below I first discuss the Hebrew experiments of Berent *et al*. (§4.1.1) and then the English experiments (§4.1.2), showing that lexical statistics had little or no effect on the response patterns in these experiments.

4.1.1 Hebrew and the OCP

Because no word usage frequency counts are available for Hebrew, Berent *et al*. could not calculate lexical neighborhood density or transitional probabilities for the tokens used in their experiments. It is therefore not possible to show definitively that these two types of lexical statistics did not influence their results. Even so, there is some evidence that is very suggestive of the fact that their results cannot be attributed to lexical statistics.

Consider first the well-formedness judgment experiments of Berent *et al*. The data discussed above in §2.2.1 come from a study by Berent and Shimron (1997). They did not report on the lexical statistics of their tokens. However, Berent *et al*. (2001a) conducted the exact same kind of well-formedness judgment experiments with the same experimental design as Berent and Shimron (1997), and attained results that were exactly comparable to those of Berent and Shimron (1997) discussed above. Berent *et al*. (2001a) did report on some lexical statistics of their tokens. I will therefore discuss the results of Berent *et al*. (2001a) here, and make the assumption that what is true about their results can be transferred to the results of Berent and Shimron (1997).

A reminder of the design and results of the experiments: Tokens consisted of forms that did not correspond to actual Hebrew words. The tokens were of three kinds – forms with no identical stem consonants ([QiSeM], no-geminate forms), forms with identical consonants in the last two stem positions ([QiSeS], final-geminate forms), and forms with identical consonants in the initial two stem positions ([QiQeS], initial-geminate forms). Based on the OT analysis developed for these forms, it was predicted that they are related as follows in terms of their well-formedness: |No-gemination ™ Final-gemination ™ Initial-gemination|. This is indeed how these forms were rated by the subjects in the experiments of Berent *et al*. What we need to show is that this rating does not correspond to the lexical statistics of these forms – i.e. that the lexical statistics did not favor no-geminate forms over final-geminate forms, and final-geminate forms over initial-geminate forms. This cannot be shown for the comparison between initial-geminate forms and the other two kinds of forms. However, it can be shown for the comparison between no-geminate forms and final-geminate forms.

Berent *et al*. (2001a) did not calculate any lexical statistic that can be correlated to lexical neighborhood density. But they did calculate a statistic that corresponds closely to transitional probability. They compiled a list all of the productive stems in the Even-Shoshan Hebrew dictionary (1993). This resulted in a list of 1,412 forms. Geminates were treated as tri-consonantal in this list – the fact that the root /S-M/ is productive in Hebrew, led to the inclusion of the form |S-M-M| in this list. They then used this list to calculate the positional type bigram frequency of their experimental forms by adding the frequency of their initial C_1C_2 bigram, their final C_2C_3 bigram, and the bigram of the initial and final consonants, C_1-C_3 (Berent *et al*., 2001a:28). In this calculation geminates were therefore treated as tri-consonantal forms. In this way a cumulative bigram frequency index can be calculated for each of the tri-consonantal forms used in their experiment. This is comparable to non-frequency weighted transitional probabilities counts (see §1.2.1).[64] The results of their calculation are shown in the table in (71).

(71) **Summed type positional bigram frequency of the forms used in the well-formedness judgment experiments by Berent *et al*. (2001a)**

Type	Example	Average summed bigram frequency
Initial-gemination	Q-Q-S	6.041
Final-gemination	Q-S-S	11.58
No-gemination	Q-S-M	9.33

[64] It would, of course, have been better if these counts could have been weighted according to the usage frequency of the items in their list of stems. However, such counts are not available for Hebrew. Also, recall the result of Bailey and Hahn (Bailey and Hahn, 1998) for transitional probabilities in English discussed in §1.2.1 – they compared frequency weighted transitional probabilities with non-frequency weighted transitional probabilities and found a very high correlation (r^2 = .96).

Berent *et al*. report that the differences in summed bigram frequencies between the three types of tokens were significant. Not surprisingly, the bigram frequency of initial-geminates was the lowest – this is expected since there are very few initial-geminates in the lexicon of Hebrew. It is therefore possible that the rejection of initial-geminates could be ascribed to the fact that these forms all had lower average summed bigram frequencies. However, the average summed bigram frequency of the final-geminate forms was higher than that of the no-geminate forms. If the bigram frequencies determined the ratings assigned to these forms, then we would have expected final-geminate forms to be rated better than no-geminate forms. The fact that the opposite was found serves as strong evidence that the ratings cannot be attributed to the effect of bigram frequencies, and by extension therefore probably not to other kinds of lexical statistics.

Now consider the lexical decision experiments of Berent *et al*. (2001b) discussed in §2.2 above. In this experiment they presented subjects with the same kinds of tokens as that used in the well-formedness judgment experiments. Based on the OT analysis developed for these forms, the prediction was that the rejection times for the different token types would be related as follows: RT(No-geminate) > RT(Final-geminate) > RT(Initial-geminate). The results of the experiment of Berent *et al*. (2001b) did not correspond exactly to this prediction.[65] For instance, they found contra the predictions that final-geminate forms were responded to slower than no-geminate forms. Even though this goes against the predictions of the analysis, we can still consider whether the response pattern can be attributed to the influence of lexical statistics. Using the same

[65] On their results and on why it deviated from the predictions, see the discussion in §2.2.2 above.

method described just above, Berent *et al.* calculated summed bigram frequencies for the no-geminate and the final-geminate forms used in their experiments. These statistics are reported in the table in (72). They found that the bigram frequency for these two kinds of tokens did not differ significantly. Therefore, the difference in response time associated with these two kinds of tokens cannot be attributed to their bigram frequencies.

(72) **Summed type positional bigram frequency of the forms used in the lexical decision experiments by Berent *et al*. (2001b)**[66]

Type	Example	Average summed bigram frequency
Final-gemination	Q-S-S	10.63
No-gemination	Q-S-M	10.93

Based on this discussion we can conclude that it is very unlikely that response patterns in the experiments of Berent *et al*. are the result of the influence of lexical statistics.

4.1.2 [sCvC]-forms in English

In the design of the experiments that I conducted on the processing of [sCvC]-forms in English I did control for the potential influence of both lexical neighborhood density and transitional probabilities on the results of the experiments. In all of the experiments I selected tokens such that these lexical statistics either did not differ between the different token kinds, or such that the lexical statistics predicted response patterns opposite to that predicted by the grammatical analysis that I developed. The fact that the results of the experiments confirmed the predictions of the grammatical analysis therefore counts as

[66] Berent *et al*. (2001b) unfortunately do not report the bigram frequency for the initial-geminate tokens that they used in their experiment.

strong evidence in favor of the claim that these results do reflect the influence of grammar rather than of lexical statistics.[67]

Consider the well-formedness judgment experiments. In these experiments subjects had to rate non-words of the form [sTvT], [sKvK] and [sPvP]. Based on the grammatical analysis that I developed for these forms (§3.1), the prediction was that these forms would be rated as follows: |sTvT ™ sKvK ™ sPvP|. The lexical statistics either favored the less well-formed tokens (for instance, favoring [sPvP] over [sKvK]), or did not differ significantly between the token types. Lexical statistics therefore either made predictions that conflicted with the grammatical analysis, or predicted no difference between token types. In spite of these lexical statistics, the experimental results show that subjects did indeed rate the forms as predicted by the grammatical analysis.

I also conducted a lexical decision experiment. The predictions here was that the reaction times to the different token types would be related as follows: RT([sTvT]) > RT([sKvK]) > RT([sPvP]). Again, lexical statistics either predicted no difference between the token types or made predictions counter to these. In spite of the lexical statistics, it was found that subjects did respond to the tokens as predicted based on the grammatical analysis.

This alone is enough to show that the results of these experiments cannot be attributed to the influence of lexical statistics. However, it is possible to show this even more conclusively. We can perform regression analyses on the response data in each experiment, using the lexical statistics as the independent variable. In this manner it is

---

[67] The details about the tokens used and the lexical statistics are discussed in the experimental design sections above, and I will not repeat these here (see §3.2 above).

possible to quantify the amount of variation in the response data that is accounted for by the lexical statistics. The table in (73) shows the results of these analyses.

(73) **Regression on response data from the three experiments on [sCvC]-tokens**

	Lexical Neighborhood Density	**Cumulative Bi-phone Probability**
Gradient well-formedness	$r^2 = 0.04$ $F(1,43) = 2.02, p > 0.05$	$r^2 = 0.02$ $F(1,43) = 0.86, p > 0.05$
Comparative well-formedness	$r^2 = 0.04$ $F(1,43) = 2.02, p > 0.05$	$r^2 = 0.02$ $F(1,43) = 0.86, p > 0.05$
Lexical decision	$r^2 = 0.18$ $F(1,20) = 4.28, p > 0.05$	$r^2 = 0.10$ $F(1,20) = 2.33, p > 0.05$

Lexical statistics therefore do not contribute significantly to the variance observed in any of the three experiments. Since the results of the experiments cannot be attributed to lexical statistics, I interpret the results as strong evidence that grammar does influence the processing of non-words.

4.2 Alternative grammatical analyses

If the results of the experiments do reflect the influence of grammar, then our theory of grammar should be able to account for these data. In the discussion above I claim that these data can be accounted for within the rank-ordering model of EVAL. In this section I will consider alternative grammatical accounts, and show that the rank-ordering model of EVAL accounts better for these data. I will first consider a classic OT grammar (Prince and Smolensky, 1993), showing it cannot account for these data (§4.2.1). This serves as motivation for the extensions to a classic OT grammar that I propose in this dissertation. After that I will consider stochastic OT grammars (Boersma, 1998, Boersma and Hayes, 2001), and OT grammars with variable constraint rankings (Anttila, 1997, Anttila and

Cho, 1998, Reynolds, 1994, Reynolds and Sheffer, 1994). These OT grammars were designed specifically to handle non-categorical data and therefore seem particularly well suited to the kind of data discussed in this chapter. However, I will show that these models face certain practical and conceptual problems that the rank-ordering model of EVAL can more easily overcome.

4.2.1 Classic OT

I argue for two extensions to the architecture of a classic OT grammar in this dissertation (Chapter 1 §1). First, I claim that EVAL can compare candidates that are not related to each other via a shared input. Secondly, I claim that EVAL imposes a harmonic ordering on the full candidate set, rather than just to distinguish the best candidate from the rest of the candidates. In this section I will show that both of these extensions are necessary to account for the results of the experiments discussed in this chapter.

Consider first the claim that EVAL can compare morphologically unrelated candidates. A classic OT grammar is a function from an input to an output so that EVAL can compare only candidate output forms that are generated by GEN for the same input. There is no direct way in which to relate morphologically unrelated forms to each other in terms of the grammar. This is of course true not only of classic OT, but of all grammars in the classic generative tradition. Since generative grammars are designed as functions that map a single input onto a single output, they cannot easily relate outputs that are derived from unrelated inputs to each other.

In grammars that allow no such inter-output comparison it is in principle not possible to account for the experimental results discussed above. All of these experiments imply comparison between different non-words that are not morphologically related to

each other. If we want to account for the response patterns in terms of grammar we need a grammar that is capable of comparing morphologically unrelated forms. In terms of an OT grammar, we need EVAL to be able to compare morphologically unrelated forms.

Now that the need for the first extension to an OT grammar has been established, what about the second extension? Is it also necessary to allow EVAL to impose a rank-ordering on the full candidate set? Let us consider what would happen if we allowed EVAL to compare morphologically unrelated forms but not to rank-order the full candidate set. EVAL can now compare a set of morphologically unrelated forms to each other (like the three kinds of tokens in the Hebrew experiments and the three kinds of tokens in the English experiments). However, EVAL will make only one distinction in these comparison sets, namely between the best form and the non-best forms. Even though EVAL can compare the three kinds of tokens used in the Hebrew experiments or the three kinds of tokens used in the English experiments, it is incapable of imposing a three-level harmonic ordering on them. This contrasts with the results of these experiments, which showed that language users do make a three level well-formedness distinction.

Since language users do make multi-level well-formedness distinctions we need a theory of grammar that can also do this. A theory that distinguishes only between the best and the rest cannot adequately account for the response patterns observed in the experiments.

In order to account for the response patterns observed in the experiments, both extensions to the classic OT model are necessary. Since language users can compare forms that are not related to each other via a shared input, EVAL should also be able to

do this. Since language users make finer distinctions than simply between the best and the non-best candidates, EVAL should also be able to make such multi-level well-formedness distinctions.

4.2.2 Stochastic OT

Stochastic OT grammars (Boersma, 1998, Boersma and Hayes, 2001) were developed to account for variable phenomena. The response data discussed in this chapter are variable. It would therefore seem that stochastic OT should be able to account for these variable response data. However, in this section I will show that these models face both conceptual and principled problems with regard to the data discussed in this chapter. I will point out three kinds of problems. The first two are of a conceptual nature, and I refer to these problems as the "which-values" and the "grammar-alone" problems (see also Chapter 5 §3.1.1). The third problem is a more principled problem, and focuses and the problems faced by these grammars with relative well-formedness difference between possible and impossible words.

The which-values problem. The rank-ordering model of EVAL that I propose in this dissertation generates information only about relative well-formedness relations between forms. Consider the English example discussed in §3. From the rank-ordering model we get the information, for instance, that [sTvT]-forms are more well-formed than [sPvP]-forms. However, no information is generated about the absolute size of the well-formedness difference between these two kinds of forms. From this follows the prediction that [sTvT]-forms will be rated better than [sPvP]-forms. But no prediction is made about how much better [sTvT]-forms will be rated. Similarly, the prediction is that [sTvT] non-words will be rejected more slowly than [sPvP] non-words. But no prediction

is made about how much the rejection times between these two kinds of tokens will differ. This represents an important difference between the rank-ordering model of EVAL and stochastic grammars. A stochastic OT grammar makes predictions about the absolute size of well-formedness differences between tokens (Hayes, 1997, 1998). I will first explain how a stochastic grammar can make predictions about absolute well-formedness differences, and then return to the problems that follow from this ability of these grammars.

In a stochastic OT grammar, constraints are ranked along a continuous ranking scale. Every constraint has a basic ranking value along this scale. The actual point where a constraint is ranked along the continuous ranking scale is not equivalent to its basic ranking value. A stochastic OT grammar includes a noise component – on every evaluation occasion a (positive or negative) random value is added to the basic ranking value of every constraint. The result of this addition determines the precise place where that constraint will be ranked along the continuous scale on the particular evaluation occasion.[68] As an example, consider the constraints *sTvT and *sPvP from the discussion in §3 above. Since [sPvP]-forms are more marked than [sTvT]-forms, the basic ranking value of *sPvP will be higher than that of *sTvT. For argument's sake, let us assume that the basic ranking value of *sPvP is 100, and the basic ranking value of *sTvT is 98. On any given evaluation occasion the actual ranking value of *sPvP will be selected from values distributed normally around 100, and the actual ranking value of *sTvT from values distributed normally around 98. Since the mean of the actual ranking values of

[68] The random value added to the basic ranking value has a normal distribution, with zero as its mean, and some arbitrarily chosen standard deviation that is set at the same value for all constraints. See Boersma and Hayes (2001).

*sPvP is higher than the mean of the actual ranking values of *sTvT, it is more likely that the actual ranking will be ||*sPvP o *sTvT|| than the other way around. But the opposite ranking is also possible – for instance, when a large enough negative noise value is added to the basic ranking value of *sPvP and a large enough positive value to the basic ranking value of *sTvT. By changing the distance between the basic ranking values of these two constraints, the likelihood of the two rankings ||*sPvP o *sTvT|| and ||*sTvT o *sPvP|| can be controlled to the finest detail.

Let us consider how such a stochastic model of OT can account for well-formedness judgments. Suppose that [sTvT]-forms are rated on average twice as good as [sPvP]-forms. The basic ranking values of *sTvT and *sPvP can be set such that the ranking ||*sPvP o *sTvT|| is twice as likely as the ranking ||*sTvT o *sPvP||. In 67% of all comparisons between an [sTvT]-form and an [sPvP]-form, these forms will be rated as follows |sTvT ™ sPvP|. In the other 33% of comparisons the opposite harmonic ordering will be imposed on the forms – i.e. |sPvP ™ sTvT|. The actual average well-formedness ratings of these forms will then correspond to the likelihood of the different rankings between the constraints.

Unlike the rank-ordering model of EVAL that I am proposing, a stochastic OT grammar can do more than make predictions about relative well-formedness judgments. It can make predictions about the absolute size of the difference in well-formedness ratings between different forms. And since the basic ranking values of constraints are distributed along a continuous scale, it is possible to model any difference in ratings by varying the distance between the basic ranking values of two constraints. In a similar way

the differences in reaction times in lexical decision experiments can be modeled very accurately in a stochastic OT grammar.

It seems that a stochastic OT grammar can account better for the well-formedness judgment and lexical decision data. However, this is only an apparent advantage of these grammars. In actuality, the fact that such a grammar makes predictions about the absolute size of the well-formedness difference between forms presents it with a problem: Which absolute values must be modeled? Consider again the data on the English [sCvC]-experiments. The tables in (74) contain a selection of the data from the well-formedness judgment experiments discussed in §3.2.

(74) **Comparison between [sTvT]-forms and [sPvP]-forms**

a. **Gradient well-formedness judgment experiment** (see §3.2.1.1)

	[sTvT]	[sPvP]
Mean score	3.65	2.41
Score ratio[69]	.60	.40

b. **Comparative well-formedness judgment experiment** (see §3.2.1.2)

	[sTvT]	[sPvP]
Mean number of preferences	22.8	6.4
Preference ratio[70]	.78	.22

The advantage of the less marked [sTvT]-forms over the more marked [sPvP]-forms is quite different in these two experiments. It is possible to model the results of the

[69] This ratio is calculated as follows: Mean score of *x* / (Mean score of *x* + Mean score of *y*).

[70] This ratio is calculated as follows: Number of *x*-preferences / (Number of *x*-preferences + Number of *y*-preferences).

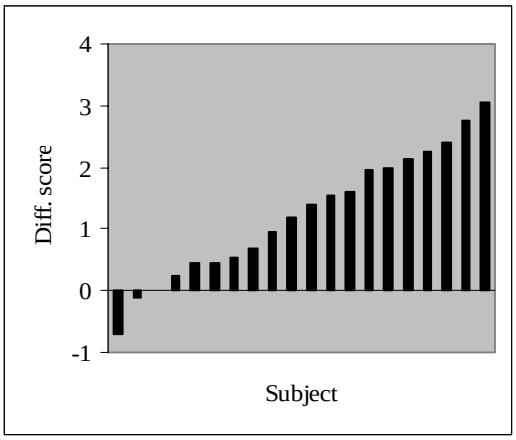
gradient well-formedness judgment experiment by placing *sKvK and *sTvT on the continuous ranking scale such that the ranking ||*sPvP o *sTvT|| will be selected 60% of the time and the ranking ||*sTvT o *sPvP|| 40% of the time. But this will then make the wrong predictions about the results of the comparative well-formedness judgment experiment. For this experiment we need the ranking ||*sPvP o *sTvT|| 79% of the time, and the ranking ||*sTvT o *sPvP|| only 21% of the time. It is not possible to model the performance in these two experiments with the same grammar. An account within a stochastic model will have to assume that the same subjects use different grammars in the different experiments.

This same problem of the stochastic models becomes evident also when we compare the differences between subjects. I will discuss the gradient well-formedness judgment experiment for English here as an example, but the same point can be made with the comparative well-formedness judgment experiment or the lexical decision experiment. In (75) I represent the difference score for the T~P-condition for each of the 20 subjects that took part in the gradient well-formedness experiment. This difference score is calculated as follows: For each subject I subtracted his/her mean rating for the [sPvP]-tokens from his/her mean rating for the [sTvT]-tokens.

The figure in (75) shows clearly that subjects differ widely in how much they rate [sTvT]-forms better than [sPvP]-forms. The ratio of a 60%-preference for [sTvT]-forms showed in the table in (74a) above is only the average across subjects. If we model our grammar based on this 60%-preference, we are therefore modeling a very abstract entity – some assumed average grammar across all of the subjects who took part in the experiment. However, it is debatable how real such an “average grammar” is. As an

alternative, we could calculate the [sTvT]-preference for every subject individually and then model the grammar for every individual subject based on his/her individual [sTvT]-preference. But this tips the scale too far in the other direction. Now the idea of some shared grammar among the subjects (as members of the same linguistic community) is lost.

(75) **Difference scores by subject for the T~P condition in the gradient well-formedness judgment experiment**



The grammar-only problem. The second conceptual problem faced by the stochastic models is closely related to the first. In the stochastic models all aspects of the variation in the performance of the subjects are modeled directly in the grammar. Implicit to this is the assumption that grammar alone is responsible for the performance of the subjects. We know that this is not the case. At least some of the variation in the ratings results not from grammar but from other factors. The difference in the results of the two well-formedness judgment experiments shows that the specific task, for instance, also contributes to the variation. And we know that factors such as lexical statistics also contribute (see discussion in §1.2.1 above). The problem with a stochastic grammar is

that it tries to account for all aspects of the variation in the data through the grammar – it does not allow enough room for the influence of other factors.

In the rank-ordering model of EVAL these problems are avoided. The rank-ordering model of EVAL predicts that grammar will rate [sTvT]-forms as better than [sPvP]-forms. However, it does not dictate the size of the [sTvT]-preferences. The actual size of the difference is the result of a complex interaction with many other variables, including task specific properties and differences between individuals. Within the rank-ordering model of EVAL the same grammar can be assumed, both for the gradient well-formedness judgment experiments and the comparative well-formedness judgment experiment. The difference in the results between these experiments can then be ascribed to factors other than grammar.

Relative well-formedness differences between possible and impossible words. [sTvT]-forms are possible words in English while [sPvP]-forms are not. In spite of this we saw that [sTvT]-forms were not rated infinitely better than [sPvP]-forms. For instance, in the comparative well-formedness judgment experiment [sPvP] was chosen over [sTvT] 22% of the time (see table (74b) above). Let us simplify by considering only the constraints *sTvT, *sPvP and IDENT[place]. The observed well-formedness difference between [sTvT]-forms and [sPvP]-forms can be captured in a stochastic grammar by assuming that we have the ranking ||*sPvP o *sTvT|| 78% of the time, and the ranking ||*sTvT o *sPvP|| 22% of the time. But this causes problems with the distinction between possible and impossible words.

English does not allow words of the form [sPvP]. This implies the following ranking between *sPvP and IDENT[place]: ||*sPvP o IDENT[place]|| (an /sPvP/-input will

then be mapped unfaithfully onto an output such as [sPvT]). There is no variation in this regard in English – an [sPvP]-form is never a possible word. The absence of variation means that the basic ranking values of *sPvP and IDENT[place] should be different enough that the ranking ||*sPvP o IDENT[place]|| will (practically) never be inverted by the addition of random noise to these basic ranking values.

But now are we predicting that variation should be observed in the way that English treats an /sTvT/-input. [sPvP] is preferred over [sTvT] 22% of the time, meaning that the ranking ||*sTvT o *sPvP|| is observed 22% of the time. /sPvP/ is never mapped faithfully, meaning that the ranking ||*sPvP o IDENT[place]|| is observed (practically) 100% of the time. This implies that the ranking ||*sTvT o IDENT[place]|| should be observed (at least) 22% of the time. And whenever this ranking is observed, then an /sTvT/-input would not be allowed to map faithfully onto itself. At least 22% of the time a word such as *state* should therefore be pronounced not as [steɪt] but as something like [steɪk] or [steɪp].[71] The rank-ordering model of EVAL avoids this problem. In this model the ranking between the constraints is always the same, namely ||*sPvP o IDENT[place] o *sTvT o Cut-off||. Because the cut-off occurs at the bottom no variation will be observed. /sTvT/ will always map faithfully onto itself and /sPvP/ will never map faithfully onto itself.

[71] I have sketched the argument here assuming that the ranking ||*sPvP o IDENT[place]|| is fixed. However, the same point can be illustrated by assuming that the ranking ||IDENT[place] o *sTvT|| is fixed (since *state* is always pronounced as [steɪt]). The only difference is then that the prediction is that we will observe the ranking ||IDENT[place] o *sPvP|| 22% of the time. We are therefore expecting that an /sPvP/-input should map faithfully onto itself at least 22% of the time. An impossible word is now predicted to be possible at least some of the time.

I have illustrated the point here with a specific example. However, this is a general problem that stochastic grammars will face whenever possible and impossible words are compared.

4.2.3 Crucially unranked constraints

Anttila (1997) proposes an extension to the classic OT grammar in order to account for non-categorical (variable) phenomena. He assumes that the constraint hierarchy for some language can contain a set of crucially unranked constraints. On every evaluation occasion one complete ranking is selected at random from amongst the different complete rankings possible between the crucially unranked constraints. Variation arises if the different rankings between these crucially unranked constraints select different candidates as optimal. This theory also makes explicit predictions about absolute frequencies. Suppose that there are n unranked constraints. There are then $n!$ possible complete rankings between these constraints. Now suppose that m of these $n!$ possible rankings select some candidate as optimal. The prediction is that this candidate would be selected as optimal $m/n!$ of the time.[72]

How could this model be applied to the well-formedness rating and lexical decision experiments discussed above? Consider the comparative well-formedness judgment experiment for English (see §3.2.1.2) as an example. In this experiment subjects had to compare *inter alia* [sTvT]-forms with [sPvP]-forms. On some of these

[72] Independently from Anttila, Reynolds proposed a very similar extension to classic OT (Reynolds, 1994). Rather than having a set of crucially unranked constraints, Reynolds assumes that there are "floating constraints". A floating constraint is a constraint that is unranked relative a span of the (ranked) constraint hierarchy. This floating constraint version of OT can be applied to well-formedness judgments and lexical decisions in the same way as Anttila's crucially unranked constraint theory. I therefore do not discuss the floating constraint theory separately.

comparisons they preferred [sTvT] over [sPvP] (78% of the time), but on other comparison occasions they preferred [sPvP] over [sTvT] (22% of the times) (see (74b) above). The well-formedness rating |sTvT ™ sPvP| would result from the ranking ||*sPvP o *sTvT||, and the rating |sPvP ™ sTvT| would result from the ranking ||*sTvT o *sPvP||. The fact that both |sTvT ™ sPvP| and |sPvP ™ sTvT| were observed can be interpreted as evidence that the constraints *sTvT and *sPvP are crucially unranked. Every time a subject in the experiment had to compare an [sPvP]-form and an [sTvT]-form, one of the two rankings between these two constraints is selected at random. If the ranking ||*sPvP o *sTvT|| is selected, then the subject prefers [sTvT] over [sPvP]. If the ranking ||*sTvT o *sPvP|| is selected, the subject prefers [sPvP] over [sTvT].[73]

This model seems capable of accounting for the variation in the responses of the subjects. However, this account faces the same three problems as the stochastic account discussed just above in §4.2.2. In addition to these, it is also faced by another more practical problem. I will discuss this practical problem first, and then briefly show why this model faces the same problem as the stochastic models.

A practical problem. In the comparison between [sTvT] and [sPvP], [sTvT] was preferred more often. Assume that *sTvT and *sPvP are indeed unranked with regard to each other. There are only two possible rankings between these two constraints. If it is indeed the case that one of the possible rankings between the unranked constraints are selected at random at every evaluation occasion, then the ranking ||*sTvT o *sPvP|| will

[73] Of course, subjects also compared [sTvT]-forms with [sKvK]-forms, and [sKvK]-forms with [sPvP]-forms. Also in these comparisons the preferences went in both directions. The implication would therefore be that all three of the constraints *sTvT, *sKvK and *sPvP are crucially unranked with respect to each other.

be selected half the time and the opposite ||*sPvP o *sTvT|| half the time. From this follows the prediction that [sTvT] will be preferred over [sPvP] half the time, and [sPvP] over [sTvT] half the time. And this is not what was observed – see that data in (74).

This problem stems from the fact that there is one constraint each against [sPvP]-forms and [sTvT]-forms. Had there been more constraints against [sPvP]-forms than against [sTvT]-forms, then it would be possible to derive the fact that [sTvT] is preferred more frequently. This problem is therefore not a general problem for the Anttila-theory. It is at least in principle possible to analyze the restriction on [sCvC]-forms such that an [sPvP]-form will violate more constraints than an [sTvT]-form.[74]

The which-value problem. Like the stochastic models, the crucially unranked constraint grammars also model absolute difference in well-formedness between different token types. Like the stochastic models, these grammars are therefore faced with the problem that there are several absolute values that measure that well-formedness difference between the token types. Different experiments show different absolute values, different individuals differ from each other. Which of these values should be modeled?

The grammar-only problem. Again like the stochastic grammars, these grammars model all variation in the response data via the grammar. This implies that grammar alone is responsible for the performance of subjects in the experiments. It does not leave enough room for others factors that are known to influence performance – factors such as lexical statistics, experimental design, individual differences, etc.

---

[74] For instance, it is possible that the *sCvC-constraints stand to each other in a stringency relation – i.e. [sPvP] violates all three of the *sCvC-constraints, [sKvK] violates *sKvK and *sTvT, and [sTvT] violates only *sTvT. See de Lacy (2002, 2003) for such an interpretation of markedness scales.

Since the rank-ordering model of EVAL models only relative well-formedness differences between token types it can avoid both of these problems. It does not have to select a specific set of absolute values. And it does not model all aspects of the variation. It only stipulates the relative well-formedness differences between tokens. The absolute differences are determined by a complex interaction of grammar with many non-grammatical factors.

Relative well-formedness differences between possible and impossible words. This exact same problem is faced by stochastic grammars, and it has to do with the fact that possible words are not rated infinitely better than impossible words. Since [sPvP] is never a possible word, the ranking ||*sPvP o IDENT[place]|| is fixed. /sPvP/ is then always mapped unfaithfully onto something like [sPvT] or [sKvP]. In the comparative well-formedness judgment experiment [sPvP] was preferred over [sTvT] 22% of the time (see table (74b) above). This means that the constraints *sPvP and *sTvT should be allowed to freely rerank. At least some of the time the ranking ||*sTvT o *sPvP|| will then be observed. But because of the fixed ranking ||*sPvP o IDENT[place]||, this means that the ranking ||*sTvT o IDENT[place]|| will also be observed at least some of the time. And from this follows the prediction that an /sTvT/-input should some times not be allowed to map faithfully onto itself. It is predicted that the word *state* should some times be pronounced as something like [steɪt] or [steɪk].[75]

[75] I have sketched the argument here assuming that the ranking ||*sPvP o IDENT[place]|| is fixed. However, the same point can be illustrated by assuming that the ranking ||IDENT[place] o *sTvT|| is fixed (since *state* is always pronounced as [steɪt]). The only difference is then that the prediction is that we will observe the ranking ||IDENT[place] o *sPvP|| some of the time. We are therefore expecting that an /sPvP/-input should map faithfully onto itself at least some of the time. An impossible word is now predicted to be possible at least some of the time.

This problem was illustrated here with a specific example. However, this is a problem that will arise every time a possible and an impossible word are compared.

The rank-ordering model of EVAL avoids all of these problems. In this model non-categorical behavior of language users does not arise from variation in the grammar (in the constraint ranking). For English the constraints are always ranked ||*sPvP o *sKvK o IDENT[place] o *sTvT o Cut-off||. Since the ranking between the constraints does not vary, IDENT[place] always outranks *sTvT so that [sTvT] is always predicted as a possible word. Similarly, both *sPvP and *sKvK always outrank IDENT[place], so that neither [sKvK] nor [sPvP] is every predicted as a possible word.

5. Concluding remarks

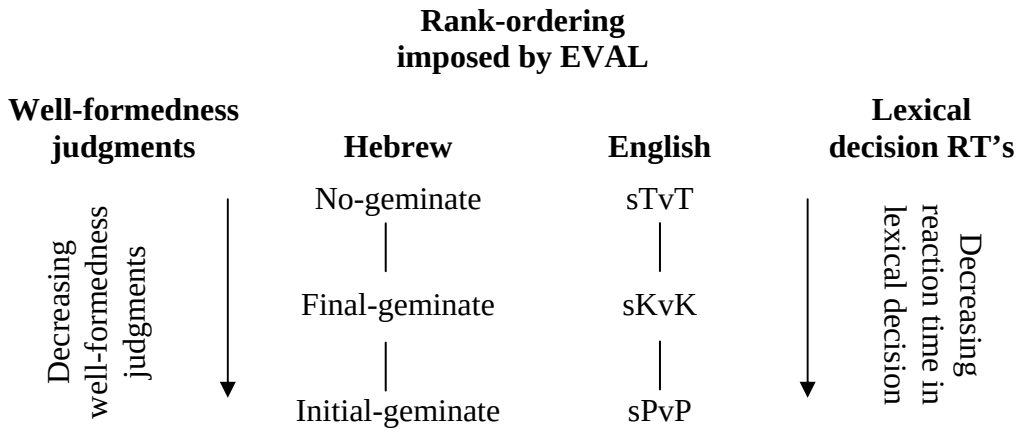
In this dissertation I argue for two extensions to a classic OT grammar: (i) First, I argue that EVAL should be allowed to compare forms that are not related to each other via a shared input. (ii) Secondly, I argue that EVAL imposes a harmonic rank-ordering on the full candidate set. In this chapter I have shown how these two extensions enable us to account for response patterns observed in well-formedness judgment experiments and lexical decision experiments.

In well-formedness judgment experiments and lexical decision experiments language users are asked to compare different non-words (either directly as in comparative well-formedness experiments, or implicitly as in gradient well-formedness judgment experiments and lexical decision experiments). Non-words used in these experiments are not related to each other via a shared input. The assumption that these

forms can be grammatically compared therefore implies that EVAL can compare such morphologically unrelated forms.

I have argued that the influence of grammar on well-formedness judgments and lexical decision can be accounted for as follows: The more well-formed a token is according to grammar (the higher slot it occupies in the rank-ordering imposed by EVAL on the candidate set), the more well-formed it will judged to be. Similarly, a non-word that is more well-formed according to grammar will be considered more seriously as potential word and will therefore be rejected more slowly as a non-word in a lexical decision task. I have illustrated this with an example on how the OCP influences the processing of non-words in Hebrew, and how a restriction on [sCvC]-forms influences the processing of non-words in English. The predictions about Hebrew and English are represented graphically in (76).

(76) **Hebrew and English in a rank-ordering model or EVAL**



The results of the experiments discussed above confirmed these predictions. This is an important result for two reasons: First, it serves as evidence for the rank-ordering model of EVAL – the predictions that follow from this model are confirmed. Secondly, it also serves as evidence for the analyses of Hebrew geminates and of English [sCvC]-

forms. The fact that the subjects in the experiments responded according to the predictions that follow from the analyses, confirms the correctness of the analyses.

The rank-ordering model of EVAL then serves two purposes: It expands the coverage of our theory in that we can now account for non-categorical phenomena such as well-formedness judgments and reaction times in lexical decision. Secondly, it adds to the kind of data that we can use as information about the grammar of some language – we can now use the results of well-formedness judgment experiments and lexical decision experiments as data to develop and test grammatical analyses.

Appendix: Tokens used in [sCvC]-experiments

A.1 Gradient well-formedness judgment experiment (see §3.2.1.1)

(77) **[sTvT] and [sKvK]-tokens**

[sTvT]	ND	Cumulative Probability	[sKvK]	ND	Cumulative Probability
stɔɪt	10.65	6.92×10^{-12}	skeɪk	26.06	2.24×10^{-9}
stɑ:t	44.16	2.56×10^{-8}	skæk	22.12	6.45×10^{-9}
stɔ:t	23.06	3.9×10^{-9}	skɑ:k	18.41	2.26×10^{-9}
stʊt	17.06	4.36×10^{-10}	skɛk	12.65	2.41×10^{-9}
stʌt	42.19	1.43×10^{-8}	skɪk	37.31	2.16×10^{-8}
	27.43	8.85×10^{-9}		23.31	6.99×10^{-9}

***t*-test on difference in ND:** $t(8) = 0.52$, two-tailed $p = 0.62$.
***t*-test on difference in Probability:** $t(8) = 0.30$, two-tailed $p = 0.77$.

(78) **[sTvT] and [sPvP]-tokens**

[sTvT]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability
stʊt	17.06	4.36×10^{-10}	spɑ:p	21.41	2.84×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	spɛp	20.32	2.40×10^{-9}
stʌt	42.19	1.43×10^{-8}	spɪp	26.69	5.48×10^{-9}
stɑ:t	44.16	2.56×10^{-8}	spæp	17.98	2.24×10^{-9}
stɔ:t	23.06	3.9×10^{-9}	spi:p	27.45	4.23×10^{-9}
	27.43	8.85×10^{-9}		22.77	3.44×10^{-9}

***t*-test on difference in ND:** $t(8) = 0.67$, two-tailed $p = 0.52$.
***t*-test on difference in Probability:** $t(8) = 1.09$, two-tailed $p = 0.31$.

(79) **[sKvK] and [sPvP]-tokens**

[sKvK]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability
skɑɪk	26.06	2.24×10^{-9}	spæp	17.98	2.24×10^{-9}
skaʊk	9.95	2.45×10^{-14}	spɑ:p	21.41	2.84×10^{-9}
ski:k	28.63	5.35×10^{-10}	spɛp	20.32	2.40×10^{-9}
skɔ:k	10.04	3.88×10^{-10}	spɪp	26.69	5.48×10^{-9}
sku:k	12.64	6.30×10^{-11}	spi:p	27.45	4.23×10^{-9}
	17.46	6.45×10^{-10}		22.77	3.44×10^{-9}

***t*-test on difference in ND:** $t(8) = 1.18$, two-tailed $p = 0.27$.
***t*-test on difference in Probability:** $t(8) = 3.76$, one-tailed $p = 0.005$.

A.2 Comparative well-formedness judgment experiments (see §3.2.1.2)

(80) [sTvT]~[sKvK]

[sTvT]	ND	Cumulative Probability	[sKvK]	ND	Cumulative Probability	Diff in ND	Diff in Cumulative Probability
stɔ:t	23.06	3.90×10^{-9}	skæk	22.12	6.45×10^{-9}	0.95	-2.55×10^{-9}
stʊt	17.06	4.36×10^{-10}	skɑ:k	18.41	2.26×10^{-9}	-1.35	-1.83×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skɛk	12.65	2.41×10^{-9}	-2.00	-2.41×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skɑ:k	18.41	2.26×10^{-9}	-7.76	-2.26×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skʌk	19.38	1.92×10^{-9}	-8.73	-1.91×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skæk	22.12	6.45×10^{-9}	-11.47	-6.44×10^{-9}
stʊt	17.06	4.36×10^{-10}	skeɪk	26.06	2.24×10^{-9}	-9.00	-1.81×10^{-9}
stʌt	42.19	1.43×10^{-8}	skɪk	37.31	2.16×10^{-8}	4.88	-7.27×10^{-9}
stɔ:t	23.06	3.90×10^{-9}	skɪk	37.31	2.16×10^{-8}	-14.25	-1.77×10^{-8}
stʊt	17.06	4.36×10^{-10}	skɪk	37.31	2.16×10^{-8}	-20.25	-2.12×10^{-8}
stʊt	17.06	4.36×10^{-10}	skæk	22.12	6.45×10^{-9}	-5.05	-6.01×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skaɪk	13.89	6.67×10^{-10}	-3.24	-6.60×10^{-10}
stɔɪt	10.65	6.92×10^{-12}	skeɪk	26.06	2.24×10^{-9}	-15.41	-2.24×10^{-9}
stʊt	17.06	4.36×10^{-10}	skɛk	12.65	2.41×10^{-9}	4.42	-1.98×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skɪk	37.31	2.16×10^{-8}	-26.66	-2.16×10^{-8}
	16.54	1.62×10^{-9}		24.21	8.14×10^{-9}		

***t*-test on difference in ND:** $t(14) = 3.32$, one-tailed $p = 0.003$.
***t*-test on difference in Probability:** $t(14) = 3.43$, one-tailed $p = 0.002$.

(81) [sTvT]~[sPvP]

[sTvT]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability	Diff in ND	Diff in Cumulative Probability
stɔ:t	23.06	3.9×10^{-9}	spɪp	26.69	5.48×10^{-9}	-3.63	-1.58×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spʌp	11.38	1.31×10^{-9}	-0.73	-1.31×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spi:p	27.45	4.23×10^{-9}	-16.80	-4.23×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spɪp	26.69	5.48×10^{-9}	-16.04	-5.47×10^{-9}
stɔ:t	23.06	3.9×10^{-9}	spi:p	27.45	4.23×10^{-9}	-4.39	-3.31×10^{-10}
stɔɪt	10.65	6.9×10^{-12}	spɑ:p	21.41	2.84×10^{-9}	-10.76	-2.83×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spæp	17.98	2.24×10^{-9}	-7.33	-2.23×10^{-9}
stʊt	17.06	4.4×10^{-10}	spæp	17.98	2.24×10^{-9}	-0.92	-1.80×10^{-9}
stʊt	17.06	4.4×10^{-10}	spɪp	26.69	5.48×10^{-9}	-9.62	-5.04×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	speɪp	17.13	8.52×10^{-10}	-6.48	-8.45×10^{-10}
stɔɪt	10.65	6.9×10^{-12}	spɛp	20.32	2.40×10^{-9}	-9.67	-2.40×10^{-9}
stʊt	17.06	4.4×10^{-10}	spɑ:p	21.41	2.84×10^{-9}	-4.35	-2.40×10^{-9}
stʊt	17.06	4.4×10^{-10}	spi:p	27.45	4.23×10^{-9}	-10.39	-3.80×10^{-9}
stʊt	17.06	4.4×10^{-10}	spɛp	20.32	2.40×10^{-9}	-3.25	-1.97×10^{-9}
stʊt	17.06	4.4×10^{-10}	spʌp	11.38	1.31×10^{-9}	5.68	-8.76×10^{-10}
	14.87	6.99×10^{-10}		21.45	3.17×10^{-9}		

***t*-test on difference in ND:** $t(14) = 4.30$, one-tailed $p < 0.000$.
***t*-test on difference in Probability:** $t(14) = 6.21$, one-tailed $p < 0.000$.

(82) **[sKvK]~[sPvP]-tokens**

[sKvK]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability	Diff in ND	Diff in Cumulative Probability
skaʊk	9.95	2.45×10^{-14}	spæp	17.98	2.24×10^{-9}	-8.03	-2.24×10^{-9}
sku:k	12.64	6.30×10^{-11}	spæp	17.98	2.24×10^{-9}	-5.34	-2.18×10^{-9}
skaʊk	9.95	2.45×10^{-14}	spɑ:p	21.41	2.84×10^{-9}	-11.46	-2.84×10^{-9}
sku:k	12.64	6.30×10^{-11}	spɑ:p	21.41	2.84×10^{-9}	-8.77	-2.78×10^{-9}
skʊk	3.86	5.76×10^{-10}	spɑ:p	21.41	2.84×10^{-9}	-17.55	-2.26×10^{-9}
skaʊk	9.95	2.45×10^{-14}	spɛp	20.32	2.4×10^{-9}	-10.37	-2.40×10^{-9}
sku:k	12.64	6.30×10^{-11}	spɛp	20.32	2.4×10^{-9}	-7.68	-2.34×10^{-9}
skʊk	3.86	5.76×10^{-10}	spɛp	20.32	2.4×10^{-9}	-16.46	-1.83×10^{-9}
ski:k	28.63	5.35×10^{-10}	spɪp	26.69	5.48×10^{-9}	1.94	-4.95×10^{-9}
skɔ:k	10.04	3.88×10^{-10}	spɪp	26.69	5.48×10^{-9}	-16.64	-5.09×10^{-9}
skʊk	3.86	5.76×10^{-10}	spɪp	26.69	5.48×10^{-9}	-22.83	-4.90×10^{-9}
skɔ:k	10.04	3.88×10^{-10}	spi:p	27.45	4.23×10^{-9}	-17.41	-3.85×10^{-9}
skʊk	3.86	5.76×10^{-10}	spi:p	27.45	4.23×10^{-9}	-23.59	-3.66×10^{-9}
skaʊk	9.95	2.45×10^{-14}	spʌp	11.38	1.31×10^{-9}	-1.43	-1.31×10^{-9}
sku:k	12.64	6.30×10^{-11}	spʌp	11.38	1.31×10^{-9}	1.26	-1.25×10^{-9}
	10.30	2.58×10^{-10}		21.26	3.18×10^{-9}		

***t*-test on difference in ND:** $t(14) = 5.26$, one-tailed $p < 0.000$.
***t*-test on difference in Probability:** $t(14) = 8.87$, one-tailed $p < 0.000$.

A.3 Lexical decision experiment (see §3.2.2)

(83) **[sTvT] and [sKvK]-tokens**

[sTvT]	ND	Cumulative Probability	[sKvK]	ND	Cumulative Probability
stɔ:t	23.06	3.90×10^{-9}	skeɪk	26.06	2.24×10^{-9}
staɪt	40.01	5.96×10^{-8}	ski:k	28.63	5.35×10^{-10}
stɛt	38.11	5.16×10^{-8}	skɪk	37.31	2.16×10^{-8}
stu:t	39.66	1.27×10^{-7}	skæk	13.89	6.45×10^{-9}
stʊt	17.06	4.36×10^{-10}	sku:k	22.12	6.30×10^{-11}
	31.58	4.85×10^{-8}		25.60	6.18×10^{-9}

***t*-test on difference in ND:** $t(8) = 0.97$, two-tailed $p = 0.36$.
***t*-test on difference in Probability:** $t(8) = 1.81$, two-tailed $p = 0.11$.

(84) **[sTvT] and [sPvP]-tokens**

[sTvT]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability
stɔːt	23.06	3.90×10^{-9}	spɑːp	21.41	2.84×10^{-9}
staɪt	40.01	5.96×10^{-8}	spaɪp	30.94	9.92×10^{-11}
stɛt	38.11	5.16×10^{-8}	spiːp	27.45	4.23×10^{-9}
stuːt	39.66	1.27×10^{-7}	spɪp	26.69	5.48×10^{-9}
stʊt	17.06	4.36×10^{-10}	spuːp	26.28	4.64×10^{-11}
	31.58	4.85×10^{-8}		26.56	2.54×10^{-9}

***t*-test on difference in ND:** $t(8) = 0.99$, two-tailed $p = 0.35$.
***t*-test on difference in Probability:** $t(8) = 1.99$, two-tailed $p = 0.08$.

(85) **[sKvK] and [sPvP]-tokens**

[sKvK]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability
skaɪk	13.89	6.67×10^{-10}	spaɪp	30.94	9.92×10^{-11}
skaʊk	9.95	2.45×10^{-14}	spaʊp	5.20	1.30×10^{-13}
skɑːk	18.41	2.26×10^{-9}	spɑːp	21.41	2.84×10^{-9}
skiːk	28.63	5.35×10^{-10}	spɔːp	2.38	9.92×10^{-13}
skʊk	3.86	5.77×10^{-10}	spʌp	11.38	1.31×10^{-9}
	14.95	8.08×10^{-10}		14.26	8.50×10^{-10}

***t*-test on difference in ND:** $t(8) = 0.10$, two-tailed $p = 0.92$.
***t*-test on difference in Probability:** $t(8) = 0.06$, two-tailed $p = 0.96$.

BIBLIOGRAPHY

Alber, Birgit. 2002. Clash, Lapse and Directionality. Ms. University of Trento.

Alderete, John. 1996. A formal theory of dissimilation, Handout of talk given at Rum.J.Clam II, May 1996.

Alderete, John. 1997. Dissimilation as local conjunction. In *NELS 27: Proceedings of the North East Linguistic Society 27*, ed. Kiyomi Kusumoto, 17-32. Amherst, MA: GLSA Publications.

Anttila, Arto. 1997. Deriving variation from grammar. In *Variation, Change and Phonological Theory*, eds. Frans Hinskens, Roeland van Hout and W. Leo Wetzels, 35-68. Amsterdam: John Benjamins.

Anttila, Arto. 2002. Morphologically conditioned phonological change. *Natural Language and Linguistic Theory* 20:1-42.

Anttila, Arto, and Cho, Young-mee Yu. 1998. Variation and change in Optimality Theory. *Lingua* 104:31-56.

Anttila, Arto, and Fong, Vivienne. 2002. Variation, Ambiguity, and Noun Classes in English. Ms. New York University and National University of Singapore.

Anttila, Arto, and Revithiadou, Anthi. 2000. Variation in allomorph selection. Paper presented at *NELS 30: Proceedings of the North East Linguistic Society 30*, Amherst.

Baayen, R., Piepenbrock, R., and Gulikers, L. 1995. *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium.

Bailey, Todd M., and Hahn, Ulrike. 1998. Determinants of wordlikeness. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, eds. Morton Ann Gernsbacher and Sharon J. Derry, 90-95. London: Lawrence Erlbaum Associates.

Bakovic, Eric. 1999. Deletion, Insertion, and Symmetrical Identity. Ms. Harvard University.

Bakovic, Eric, and Wilson, Colin. 2000. Transparency, strict locality, and targeted constraints. In *WCCFL 19: The Proceedings of the West Coast Conference on Formal Linguistics XIX*, eds. Roger Billerey and Brook Danielle Lillehaugen, 43-56. Somerville, MA: Cascadilla Press.

Balota, David A., and Chumbley, James I. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance* 10:340-357.

Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language & Linguistic Theory* 12:571-596.

Bayley, Robert. 1995. Consonant cluster reduction in Tejano English. *Language Variation and Change* 6:303-326.

Bayley, Robert. 1997. Variation in Tejano English: evidence for variable lexical phonology. In *Language Variety in South Revisited*, eds. Cynthia Bernstein, Thomas Nunnally and Robin Sabino, 197-209. Tuscaloosa: The University of Alabama Press.

Berent, Iris, Everett, Daniel L., and Shimron, Joseph. 2001a. Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42:1-60.

Berent, Iris, Marcus, Gary F., Shimron, Joseph, and Gafos, Adamantios I. 2002. The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition* 83:113-139.

Berent, Iris, and Shimron, Joseph. 1997. The representation of Hebrew words: evidence from the obligatory contour principle. *Cognition* 64:39-72.

Berent, Iris, Shimron, Joseph, and Vaknin, Vered. 2001b. Phonological constraints on reading: evidence from the Obligatory Contour Principle. *Journal of Memory and Language* 44:644-665.

Berkley, Deborah Milam. 1994a. Variability and Obligatory Contour Principle effects. In *CLS 30: Proceedings of the 30th Annual Meeting of the Chicago Linguistic Society. Volume 2: The Parasession on Variation and Linguistic Theory.* Chicago: Chicago Linguistic Society.

Berkley, Deborah Milam. 1994b. The OCP and gradient data. *Studies in the Linguistic Sciences* 24:59-72.

Berkley, Deborah Milam. 2000. *Gradient Obligatory Contour Principle Effects*. Ph.D. dissertation, North Western University, Chicago.

Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. Ph.D. dissertation, University of Amsterdam.

Boersma, Paul, and Hayes, Bruce P. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.

Broselow, Ellen. 1991. The Structure of Fricative-Stop Onsets. Ms. State University of New York, Stony Brook.

Browman, Catherine P., and Goldstein, Louis. 1990. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, eds. John Kingston and Mary E. Beckman. Cambridge: Cambridge University Press.

Brown, Roger W., and Hildum, Donald C. 1956. Expectancy and the perception of syllables. *Language* 32:411-419.

Browne, Wayles. 1981. Slavic -ba and English *slil: two persistent constraints. *Folia Slavica* 4:219-225.

Carnap, Rudolf. 1962. *Logical Foundations of Probability. 2nd edition.* Chicago: University of Chicago Press.

Cedergren, Henrietta J., and Sankoff, David. 1974. Variable rules: performance as a statistical reflection of competence. *Language* 50:333-355.

Cedergren, Henrietta J., and Simoneau, L. 1985. La chute des voyelles fermées en français des Montréal: as-tu entendu la belle syncope? In *Les Tendances Dynamiques du Français à Montréal*, eds. Henrietta J. Cedergren and Monique Lemieux, 55-144. Montréal: Office de la Langue Française.

Celdran, E.M., and Villalba, X. 1995. Locus equations as a metric for place of articulation in automatic speech recognition. In *ICPhS XIII: Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden. Volume 1*, 30-33.

Chomsky, Noam, and Halle, Morris. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Clements, George N. 1988. The role of the sonority cycle in core syllabification. *Working Papers of the Cornell Phonetics Laboratory* 2:1-68.

Clements, George N., and Keyser, Samuel Jay. 1983. *CV Phonology: a Generative Theory of the Syllable*. Cambridge: MIT Press.

Coetzee, Andries W. 1996a. The Phonology of Tiberian Hebrew Words Ending on Underlying Consonant Clusters, Department of Semitic Languages, Potchefstroom University for Christian Higher Education, South Africa: M.A.

Coetzee, Andries W. 1996b. The neutral reduced vowel, vowel epenthesis and segholate formation in Tiberian Hebrew. *South African Journal of Linguistics* Supplement 31:5-21.

Coetzee, Andries W. 1999a. *Tiberian Hebrew Phonology: Focussing on Consonant Clusters*. Studia Semitica Neerlandica. Assen: Van Gorcum.

Coetzee, Andries W. 1999b. Hebrew and Aramaic segholation and the generality and ordering of phonological rules. *Journal of Semitic Studies* 44:215-226.

Coetzee, Andries W. to appear. The Obligatory Contour Principle in the Perception of English. In *Proceedings of the First Conference on Phonetics and Phonology in Iberia*, eds. Sonia Frota, Marina Vigario and Maria Joao Freitas. New York: Mouton.

Coleman, John, and Pierrehumbert, Janet. 1997. Stochastic phonological grammars and acceptability. In *Computational Phonology. Third Meeting of the ACL Special Interest Group in Computational Phonology*, 49-56. Somerset: Association for Computational Linguistics.

Crosswhite, Katherine. 2000a. The analysis of extreme vowel reduction. In *UCLA Working Papers in Linguistics 4: Papers in Phonology*, eds. Adam Albright and Taehong Cho, 1-12. Los Angeles: Department of Linguistics, UCLA.

Crosswhite, Katherine. 2000b. Sonority-driven reduction. In *BLS 26*.

Crosswhite, Katherine. 2001. *Vowel Reduction in Optimality Theory*. New York: Routledge.

Cutler, Anne, Mehler, Jacques, Norris, Dennis, and Segui. 1987. Phoneme identification and the lexicon. *Cognitive Psychology* 19:141-177.

Davey, B. A., and Priestley, H. A. 1990. *Introduction to Lattices and Order*. Cambridge: Cambridge University Press.

Davis, Stuart. 1982. On the internal organization of syllable constituents. In *Coyote Papers 3: Working Papers in Linguistics of the University of Arizona*, ed. Thomas G. Larson, 3-22. Tucson: University of Arizona.

Davis, Stuart. 1984. Some implications of onset-coda constraints for syllable phonology. Paper presented at *CLS 20: Papers from the Twentieth Regional Meeting of the Chicago Linguistics Society*, Chicago.

Davis, Stuart. 1988a. Syllable onsets as a factor in stress rules. *Phonology* 5:1-19.

Davis, Stuart. 1988b. *Topics in Syllable Geometry*. Outstanding Dissertation in Linguistics. New York: Garland.

Davis, Stuart. 1989. Cross-vowel phonotactic constraints. *Computational Linguistics* 15:109-110.

Davis, Stuart. 1991. Coronals and the phonotactics of nonadjacent consonants in English. In *The Special Status of Coronals: Internal and External Evidence*, eds. C. Paradis and J.F. Prunet, 49-60. San Diego: Academic Press.

de Lacy, Paul. 2002. *The Formal Expression of Markedness*. Ph.D. dissertation, University of Massachusetts, Amherst.

de Lacy, Paul. 2003. Conflation and hierarchies. In *Cambridge Occasional Papers in Linguistics 1*, eds. Lluisa Astruc-Aguilera and Marc Richards. Cambridge: Linguistics Department, University of Cambridge.

Diehl, Randy L., and Walsh, Margaret A. 1989. An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America* 85:2154-2164.

Eek, A., and Meister, E. 1995. The perception of stop consonants: locus equations and spectral integration. In *ICPhS XIII: Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden. Volume 1*, 18-21.

Eimas, P.D., Marcovitz Hornstein, S.B., and Payton, P. 1990. Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language* 29:160-180.

Eimas, P.D., and Nygaard, L.C. 1992. Contextual coherence and attention in phoneme monitoring. *Journal of Memory and Language* 31:375-395.

Eisner, Jason. 1997a. Efficient generation in Primitive Optimality Theory. Ms. Department of Computer and Information Science, University of Pennsylvania, Philadelphia.

Eisner, Jason. 1997b. What constraints should OT allow? Handout Presentation Given at LSA Meeting in Chicago, 4 January 1997.

Eisner, Jason. 1999. Doing OT in a straightjacket. Handout for Talk Presented at UCLA, 14 June 1999.

Elenbaas, Nine, and Kager, René. 1999. Ternary rhythm and the Lapse constraint. *Phonology* 16:273-330.

Enderton, Herbert B. 1977. *Elements of Set Theory*. New York: Academic Press.

Epstein, Samuel David. 1990. Differentiation and reduction in syntactic theory: a case theory. *Natural Language & Linguistic Theory* 8:313-323.

Epstein, Samuel David. 2000. Differentiation and reduction in syntactic theory: a case study. In *Essays in Syntactic Theory*, ed. Samuel David Epstein, 83-93. New York: Routledge.

Even-Shoshan, A. 1993. *Ha'milon ha'xadash. [Title in English: The New Dictionary.]*. Jerusalem: Kiryat-Sefer.

Fails, Willis C., and Clegg, J. Halvor. 1992. A spectrographic analysis of Portuguese stressed and unstressed vowels. In *Romance Linguistics: The Portuguese Context*, eds. Dale A. Koike and Donaldo P. Macedo, 31-42. Westport: Bergin & Garvey.

Fasold, R. 1972. *Tense Marking in Black English*. Arlington: Center for Applied Linguistics.

Firth, J.R. 1935. The use and distribution of certain English sounds. *English Studies: A Journal of English Letters and Philology* 17:8-18.

Firth, J.R. 1935-1937. Alphabets and phonology in India and Burma. *Bulletin of the School of Oriental Research (University of London)* 8:517-546.

Fleischhacker, Heidi. 2001. Cluster dependent epenthesis asymmetries. In *UCLA Working Papers in Linguistics 7: Papers in Phonology 5*, eds. Adam Albright and Taehong Cho, 71-116. Los Angeles: UCLA.

Fleischhacker, Heidi. 2002. Onset transfer in reduplication. Ms., *Paper presented at the annual meeting of the LSA, San Francisco, January 2002*. UCLA, Los Angeles. (Available at www.linguistics.ucla.edu/people/grads/fleischhacker).

Fowler, Carol A. 1994. Invariants, specifiers, cues: an investigation of locus equations as information for place of articulation. *Perception and Psychophysics* 55:597-610.

Frauenfelder, U.H, and Seguí, J. 1989. Phoneme monitoring and lexical processing: evidence for associative context effects. *Memory and Cognition* 17:134-140.

Frisch, Stefan A., Large, Nathan R., Zawaydeh, Bushra Adnan, and Pisoni, David B. 2001. Emergent phonotactic generalizations in English and Arabic. In *Frequency and the Emergence of Linguistic Structure*, eds. Joan L. Bybee and Paul Hopper, 159-179. Amsterdam: John Benjamins.

Frisch, Stefan A., Pierrehumbert, Janet B., and Broe, Michael B. 2004. Similarity Avoidance and the OCP. *Natural Language & Linguistic Theory* 22(1):179-228.

Frisch, Stefan A., and Zawaydeh, Bushra Adnan. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77:91-106.

Fruchter, David, and Sussman, Harvey M. 1997. The perceptual relevance of locus equations. *Journal of the Acoustical Society of America* 102:2997-3008.

Fudge, E.C. 1969. Syllables. *Journal of Linguistics* 5:253-286.

Fujimura, Osamu, and Lovins, Julie B. 1978. Syllables as concatenative phonetic units. In *Syllables and Segments*, eds. Alan Bell and Joan B. Hooper, 107-120. Amsterdam: North-Holland Publishing Company.

Gafos, Adamantios I. 1998. Eliminating long-distance consonantal spreading. *Natural Language and Linguistic Theory* 16:223.

Gafos, Adamantios I. 2003. Greenberg's asymmetry in Arabic: a consequence of stems in paradigms. *Language* 79:317-355.

Ganong, W.F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6:110-125.

Garr, W.R. 1989. The *seghol* and segholation in Hebrew. *Journal of Near Eastern Studies* 48:109-116.

Gnanadesikan, Amalia. 1996. Child phonology in Optimality Theory: Ranking markedness and faithfulness constraints. In *Proceedings of the 20th Annual Boston University Conference on Language Development*, eds. A. Stringfellow, D. Cahana-Amitay, E. Hughes and A. Zukowski, 237-248. Somerville, MA: Cascadilla Press.

Goldsmith, John. 1976. An overview of autosegmental phonology. *Linguistic Analysis* 2:23-68.

Gouskova, Maria. 2001. Falling sonority onsets, loanwords, and syllable contact. In *CLS 37: Papers from the 37th Meeting of the Chicago Linguistic Society. Vol. 1: The Main Session*, eds. Mary Andronis, Christopher Ball, Heidi Elston and Sylvain Neuvel, 175-185. Chicago: CLS.

Gouskova, Maria. 2003. *Deriving Economy: Syncope in Optimality Theory*. Ph.D. dissertation, University of Massachusetts, Amherst.

Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6:162-181.

Guy, Gregory R. 1980. Variation in the group and the individual. In *Locating Language in Time and Space*, ed. William Labov, 1-36. New York: Academic.

Guy, Gregory R. 1981. Parallel variability in American dialects of Spanish and Portuguese. In *Variation Omnibus*, eds. David Sankoff and Henrietta J. Cedergren, 85-96. Edmonton: Linguistics Research, Inc.

Guy, Gregory R. 1991a. Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change* 3:1-22.

Guy, Gregory R. 1991b. Contextual conditioning in variable lexical phonology. *Language Variation and Change* 3:223-239.

Guy, Gregory R. 1994. The phonology of variation. In *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society. Volume 2: The Parasession on Variation in Linguistic Theory*, eds. Katharine Beals *et al.*, 133-149. Chicago: Chicago Linguistic Society.

Guy, Gregory R. 1997. Violable is variable: Optimality Theory and linguistic variation. *Language Variation and Change* 9:333-347.

Guy, Gregory R., and Boberg, Charles. 1997. Inherent variability and the obligatory contour principle. *Language Variation and Change* 9:149-164.

Guy, Gregory R., and Boyd, Sally. 1990. The development of a morphological class. *Language Variation and Change* 2:1-18.

Hammond, Michael. 1994. An OT account of variability in Walmatjari stress. Ms. University of Arizona.

Hammond, Michael. 1997. Parsing syllables: Modeling OT computationally. Ms.

Hammond, Michael. 2000. The logic of Optimality Theory. Ms. University of Arizona.

Hartkemeyer, Dale C. 2000. **V: An Optimality Theoretic Analysis of Vowel Loss Phenomena with Special Reference to Latin, Early Western Romance, and Basque, Linguistics*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Hayes, Bruce P. 1980. *A Metrical Theory of Stress Rules*. Ph.D. dissertation, MIT, Cambridge.

Hayes, Bruce P. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13:227-276.

Hayes, Bruce P. 1985. *A Metrical Theory of Stress Rules*. Outstanding Dissertations in Linguistics. New York: Garland.

Hayes, Bruce P. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hayes, Bruce P. 1997. Gradient Well-Formedness in Optimality Theory. In *Handout for talk given at the University of Oregon on April 4, 1997*.

Hayes, Bruce P. 1998. Gradient Well-Formedness in Optimality Theory. Ms. UCLA.

Hayes, Bruce P., and MacEachern, Margaret. 1996. Folk Verse Form in English. Ms. UCLA.

Hayes, Bruce P., and MacEachern, Margaret. 1997. Quatrain Form in English Folk Verse. Ms. UCLA.

Haywood, J.A., and Nahmad, H.M. 1965. *A New Arabic Grammar of the Written Language*. London: Lund Humphries.

Hinskens, Frans. 1992. *Dialect Leveling in Limburg. Structural and Sociolinguistic Aspects*. Ph.D. dissertation, Katholieke Universiteit te Nijmegen.

Hinskens, Frans. 1996. *Dialect Leveling in Limburg: Structural and Sociolinguistic Aspects*. Tübingen: Niemeyer.

Hinskens, Frans, van Hout, Roeland, and Wetzels, W. Leo. 1997. Balancing data and theory in the study of phonological variation and change. In *Variation, Change and Phonological Theory*, eds. Frans Hinskens, Roeland van Hout and W. Leo Wetzels, 1-33. Amsterdam: John Benjamins.

Hockett, Charles F. 1955. *A Manual of Phonology*. Baltimore: Waverly Press.

Holmes, Janet. 1995. Glottal stops in New Zealand English: an analysis of variants of word-final /t/. *Linguistics* 33:433-463.

Itô, Junko, and Mester, Armin. 1997. Sympathy Theory and German truncations. In *University of Maryland Working Papers in Linguistics 5. Selected Papers from the Hopkins Optimality Workshop 1997/University of Maryland Mayfest 1997*, eds. Viola Miglio and B. Moreen. College Park: University of Maryland.

Itô, Junko, and Mester, Armin. 1998. Markedness and Word Structure: OCP effects in Japanese. Ms. University of California, Santa Cruz.

Itô, Junko, and Mester, Armin. 1999. The phonological lexicon. In *A Handbook of Japanese Linguistics*, ed. N. Tsujimura, 62-100. Oxford: Blackwell.

Itô, Junko, and Mester, Armin. 2003. On the sources of opacity in OT. In *The Syllable in Optimality Theory*, eds. Caroline Féry and Ruben van de Vijver, p. 271-303. Cambridge: Cambridge University Press.

Jakobson, Roman. 1968. *Child Language, Aphasia, and Phonological Universals.* The Hague: Mouton.

Johansonn, Stig. 1973. Linking and intrusive /r/ in English: a case of more concrete phonology. *Studia Linguistica* 27:53-68.

Kager, René. 2001. Rhythmic Directionality by Positional Licensing. Ms., *Fifth HIL Phonology Conference, University of Potsdam, 11 January 2001*.

Kahn, Daniel. 1976. *Syllable Based Generalizations in English Phonology*. Ph.D. dissertation, MIT, Cambridge.

Kahn, Daniel. 1980. *Syllable-Based Generalizations in English Phonology*. Outstanding Dissertations in Linguistics. New York: Garland.

Kang, Hyeon-Seok. 1994. Variation in past-marking and the question of the system in Trinidadian English. In *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society. Volume 2: The Parasession on Variation in Linguistic Theory*, eds. Katharine Beals *et al.*, 150-164. Chicago: Chicago Linguistic Society.

Karinš, A. Krišjanis. 1995a. Vowel deletion in Latvian. *Language Variation and Change* 7:15-34.

Karinš, A. Krišjanis. 1995b. Phonetic evidence for phonological structure: word stress in Latvian. In *Proceedings of the 13th International Congress of Phonetic Sciences*, eds. K. Elenius and P. Branderud, 642-645. Stockholm: KTH and Stockholm University.

Karttunen, Lauri. 1998. The proper treatment of optimality in computational phonology. Paper presented at *Proceedings of FSMNLP'98. International Workshop on Finite-State Methods in Natural Language Processing, June 29-July 1, 1998*, Bilkent University, Ankara, Turkey.

Kay, Paul, and McDaniel, Chad K. 1979. On the logic of variable rules. *Language in Society* 8:151-187.

Kaye, Jonathan, Lowenstamm, Jean, and Vergnaud, Jean-Roger. 1990. Constituent structure and government in phonology. *Phonology* 7:193-231.

Kenstowicz, Michael, and Green, Tom. 1995. The Lapse Constraint. Ms. Massachusetts Institute of Technology, Cambridge.

Kewley-Port, Diane. 1983. Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America* 73:322-335.

Kewley-Port, Diane, Pisoni, David B., and Studdert-Kennedy, Michael. 1983. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America* 73:1779-1793.

Kingston, John. to appear. Keeping and losing contrasts. In *BLS 28: Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*.

Kingston, John, and Shinya, Takahito. 2003. Markedness asymmetries in place perception in consonant clusters. In *ICPhS XV: Proceedings of the XVth International Congress of Phonetic Sciences, Barcelona, Spain*, 399-492.

Kiparsky, Paul. 1993. Variable Rules. In *Handout for a talk given at the First Rutgers Optimality Workshop, held at Rutgers University in October 1993.*

Kornai, András. 1995. *Formal Phonology*. New York: Garland.

Kuryłowicz, Jerzy. 1971. A problem of German alliteration. In *Studies in Language and Literature in Honour of Margaret Schlauch*, eds. Mieczyslaw Brahmer, Stanislaw Helsztynski and Julian Kryzanowski, 195-201. New York: Russell & Russell.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1989. The child as linguistic historian. *Language Variation and Change* 1:85-97.

Labov, William. 1997. Resyllabification. In *Variation, Change and Phonological Theory*, eds. Frans Hinskens, Roeland van Hout and W. Leo Wetzels, 145-179. Amsterdam: John Benjamins.

Labov, William, Cohen, P., Robins, C., and Lewis, J. 1968. *A Study of the Non-standard English of Negro and Puerto Rican Speakers in New York City. Volume 1*. Philadelphia: U.S. Regional Survey.

Ladefoged, Peter, and Maddieson, Ian. 1996. *The Sounds of the World's Languages*. Phonological theory. Oxford: Blackwell.

Lahiri, Aditi, Gewirth, Letitia, and Blumstein, Sheila E. 1984. A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study. *Journal of the Acoustical Society of America* 76:391-404.

Lamontagne, Greg. 1993. *Syllabification and Consonant Cooccurrence Conditions*. Ph.D. dissertation, University of Massachusetts, Amherst.

Leben, William. 1973. *Suprasegmental Phonology*. Ph.D. dissertation, MIT, Cambridge.

Lisker, L. 1986. <<Voicing>> in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29:3-11.

Lisker, L., and Abramson, A. 1964. A cross linguistic study of voicing in initial stops: acoustical measurements. *Word* 20:384-422.

Lombardi, Linda. 1998. Evidence for MaxFeature constraints from Japanese. *University of Maryland Working Papers in Linguistics* 7:41-64.

Lombardi, Linda. 2001. Why Place and Voice are different: constraint-specific alternations in Optimality Theory. In *Segmental Phonology in Optimality Theory. Constraints and Representations*, ed. Linda Lombardi, 13-45. Cambridge: Cambridge University Press.

Łubowicz, Anna. 1998. Derived Environment in Optimality Theory. Ms. University of Massachusetts, Amherst.

Łubowicz, Anna. 1999. Derived environment effects in OT. In *WCCFL 17: The Proceedings of the West Coast Conference on Formal Linguistics 17*, eds. Kimary N. Shahin, Susan J. Blake and Eun-Sook Kim, 451-465. Stanford: CSLI Publications.

Łubowicz, Anna. 2002. Derived environment effects in Optimality Theory. *Lingua* 112:243-280.

Luce, Paul A. 1986. Neighborhoods of words in the mental lexicon, Department of Psychology, Indiana University.

Luce, Paul A., and Pisoni, David B. 1998. Recognizing spoken words: the neighborhood activation model. *Ear and Hearing* 19:1-35.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge studies in speech science and communication. Cambridge: Cambridge University Press.

Major, Roy C. 1981. Stress-timing in Brazilian Portuguese. *Journal of Phonetics* 9:343-351.

Major, Roy C. 1985. Stress and rhythm in Brazilian Portuguese. *Language* 61:259-282.

Major, Roy C. 1992. Stress and rhythm in Brazilian Portuguese. In *Romance Linguistics: The Portuguese Context*, eds. Dale A. Koike and Donaldo P. Macedo, 3-30. Westport: Bergin & Garvey.

Malécot, André. 1958. The role of releases in identification of released final stops. *Language* 34:370-380.

Malone, Joseph L. 1972. A Hebrew flip-flop rule and its historical origins. *Lingua* 30:422-448.

Malone, Joseph L. 1993. *Tiberian Hebrew Phonology*. Winona Lake: Eisenbrauns.

Martinez-Gil, Fernando. 1997. Word-final epenthesis in Galician. In *Issues in the Phonology and Morphology of the Major Iberian Languages*, eds. Fernando Martinez-Gil and Alfonso Morales-Front. Washington, D.C.: Georgetown University Press.

Martins, Maria Raquel Delgado. 1964 and 1973. Analise acústica das vogais orais tónicas em portugûes. *Boletim de Filologia* 22:303-314.

Mateus, Maria Helena, and d'Andrade, Ernesto. 2000. *The Phonology of Portuguese*. The Phonology of the World's Languages. Oxford: Oxford University Press.

McCarthy, John J. 1979. Formal Problems in Semitic Phonology and Morphology, Linguistics, MIT.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12:373-418.

McCarthy, John J. 1986. OCP effects: gemination and antigemination. *Linguistic Inquiry* 17:207-263.

McCarthy, John J. 1991. Synchronic rule inversion. In *BLS 17: Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*, eds. Laurel A. Sutton, Christopher Johnson and Ruth Shields, 192-207. Berkeley: Berkeley Linguistics Society.

McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* 38:169-195.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology 3*, ed. Patricia Keating, 191-233. Cambridge: Cambridge University Press.

McCarthy, John J. 2002a. Comparative markedness. In *UMOP 26: University of Massachusetts Occasional Papers in Linguistics 26. Papers in Optimality Theory II*, eds. Angela C. Carpenter, Andries W. Coetzee and Paul De Lacy, 171-246. Amherst: GLSA.

McCarthy, John J. 2002b. *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J., and Prince, Alan. 1990a. Foot and word in prosodic morphology: the Arabic broken plural. *Natural Language & Linguistic Theory* 8:209-283.

McCarthy, John J., and Prince, Alan. 1990b. Prosodic morphology and templatic morphology. In *Perspectives on Arabic Linguistics II: Papers From the Second Annual Symposium on Arabic Linguistics*, eds. Mushira Eid and John J. McCarthy, 1-54. Amsterdam: John Benjamins.

McCarthy, John J., and Prince, Alan. 1993a. Prosodic Morphology I: Constraint Interaction and Satisfaction. Report no. RuCCS-TR-3. Ms. New Brunswick, Rutgers University Center for Cognitive Science.

McCarthy, John J., and Prince, Alan. 1993b. Generalized alignment. In *Yearbook of Morphology*, eds. Geert Booij and Jaap van Marle, 79-153. Dordrecht: Kluwer.

McCarthy, John J., and Prince, Alan S. 1994. The emergence of the unmarked: optimality in prosodic morphology. In *NELS 24: Proceedings of the North East Linguistic Society*, ed. M. Gonzales, 333-379. Amherst: GLSA, University of Massachusetts.

McCarthy, John J., and Prince, Alan S. 1995. Faithfulness and reduplicative identity. In *UMOP 18: University of Massachusetts Occasional Papers in Linguistics. Papers in Optimality Theory*, eds. Jill Beckman, Laura Walsh Dickey and Suzanne Unbanczyk, 249-384. Amherst: GLSA, University of Massachusetts.

Mees, I. 1987. Glottal stop as a prestigious feature in Cardiff English. *Englishes World-Wide* 8:25-39.

Meyer, Kuno. 1909. *A Primer of Irish Metrics. With a Glossary, and an Appendix Containing an Alphabetical List of the Poets of Ireland*. Dublin: School of Irish Learning.

Moreton, Elliott. 1999. Non-computable functions in Optimality Theory. Ms. University of Massachusetts, Amherst.

Moreton, Elliott. 2002a. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84:55-71.

Moreton, Elliott. 2002b. *Phonological Grammar in Speech Perception*. Ph.D. dissertation, University of Massachusetts, Amherst.

Moreton, Elliott. 2003. Evidence for phonological grammar in speech perception. In *Handout for the Workshop op Markedness and the Lexicon held at MIT in January 2003*.

Morgenbrod, M., and Serifi, E. 1981. The sound structure of verb roots in modern Hebrew. *Journal of Linguistics* 17:11-16.

Moscati, Sabatino. 1964. *An Introduction to the Comparative Grammar of the Semitic Languages: Phonology and Morphology*. Porta Linguarum Orientalium. Wiesbaden: Otto Harrassowitz.

Murphy, Gerard. 1961. *Early Irish Metrics*. Dublin: Royal Irish Academy.

Nagy, Naomi, and Reynolds, Bill. 1997. Optimality Theory and variable word-final deletion in Faetar. *Language Variation and Change* 9:37-55.

Nearey, T.M., and Shammass, S.E. 1987. Formant transitions as partly distinctive invariant properties in the identification of voiced stops. *Canadian Acoustics* 15:17-24.

Neu, Helene. 1980. Ranking of constraints on /t,d/ deletion in American English: a statistical analysis. In *Locating Language in Time and Space*, ed. William Labov, 37-54. New York: Academic Press.

Newman, Rochelle S., Sawusch, James R., and Luce, Paul A. 1997. Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology* 23:873-889.

Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In *Papers in Laboratory Phonology I*, eds. J. Kingston and M.E. Beckman, 258-275. Cambridge: Cambridge University Press.

O'Leary, De Lacy. 1969. *Comparative Grammar of the Semitic Languages*. Amsterdam: Philo Press.

Oliviera, Jura D. 1993. *Conversational Portuguese*. New York: Crown.

Padgett, Jaye. 1991. OCP subsidiary features. In *NELS 22: Proceedings of the Northeast Linguistic Society*, 335-346. Amherst, MA: GLSA.

Paradis, Carole, and Prunet, Jean-François. 1990. Coronal or velar placeless consonants? In *WCCFL 9: Proceedings of the West Coast Conference in Formal Linguistics 9*, ed. Aaron L. Halpern, 95-119. Stanford: Stanford Linguistics Association.

Paradis, Carole, and Prunet, Jean-François eds. 1991. *Phonetics and Phonology: The Special Status of Coronals: Internal and External Evidence*. San Diego: Academic Press.

Parker, Steve. 2002. Quantifying the Sonority Hierarchy, University of Massachusetts Amherst. [Published by the GLSA.].

Parkinson, Stephen. 1988. Portuguese. In *The Romance Languages*, eds. Martin Harris and Nigel Vincent, 131-169. New York: Oxford University Press.

Partee, Barbara H., ter Meulen, Alice, and Wall, Robert E. 1993. *Mathematical Methods in Linguistics*. Dordrecht/Boston/London: Kluwer.

Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. In *The Prosody Morphology Interface*, eds. Harry van der Hulst, René Kager and Wim Zonneveld, 310-343. Cambridge: Cambridge University Press.

Pater, Joe. 2000. Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* 17:237-274.

Patrick, Peter L. 1991. Creoles at the intersection of variable processes: *-t, d* deletion and past-marking in the Jamaican mesolect. *Language Variation and Change* 3:171-189.

Pettersson, Thore, and Wood, Sidney. 1987. Vowel reduction in Bulgarian: the phonetic data and model experiments. *Folia Linguistica* XXII:239-262.

Pierrehumbert, Janet. 1993. Dissimilarity in the Arabic verbal roots. In *NELS 23: Proceedings of the North East Linguistic Society*, ed. A. Schafer, 367-381. Amherst: GLSA.

Pierrehumbert, Janet. 1994. Syllable structure and word structure: a study of triconsonantal clusters in English. In *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, ed. Patricia A. Keating, 168-187. Cambridge: Cambridge University Press.

Pierrehumbert, Janet, Hay, Jennifer, and Beckman, Mary. In press. Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in Laboratory Phonology VI*, eds. J. Local, R. Ogden and R. Temple. Cambridge: Cambridge University Press.

Pitt, M.A., and Samuel, A.G. 1993. An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance* 19:699-725.

Prince, Alan. 1997. Harmonic Completeness, AP Order, Chain Shifts. Ms. Lecture notes for course taught at the LSA Institute in 1997.

Prince, Alan. 1998. Two lectures in Optimality Theory. Ms. Lectures given at the Kobe University in September 1998.

Prince, Alan. 2002. Entailed ranking arguments. Ms. Rutgers University.

Prince, Alan, and Smolensky, Paul. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Prince, Alan, and Smolensky, Paul. 1997. Optimality: From neural networks to Universal Grammar [14 March 1997]. *Science* 275:1604-1610.

Pullam, Geoffrey. 1976. The Duke of York Gambit. *Journal of Linguistics* 12:83-102.

Repp, Bruno H. 1978. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception and Psychophysics* 24:471-485.

Repp, Bruno H. 1982. Perceptual assessment of coarticulation in sequences of two stop consonants. In *Status Report on Speech Research. SR-71/72*. New Haven: Haskins Laboratories.

Repp, Bruno H. 1983. Bidirectional contrast effects in the perception of VC-CV sequences. *Perception and Psychophysics* 33:147-155.

Reynolds, Bill. 1994. *Variation and Phonological Theory*. Ph.D. dissertation, University of Pennsylvania.

Reynolds, Bill, and Sheffer, Haddas. 1994. Variation and optimality. In *University of Pennsylvania Working Papers in Linguistics. Volume 1.*, eds. Roumyana Izvorski, Miriam Meyerhoff, Bill Reynolds and Victoria Tredinnick, 103-110. Philadelphia: Department of Linguistics, University of Pennsylvania.

Rogers, Francis Millet. 1948. Insular Portuguese pronunciation: Porto Santo and the eastern Azores. *Hispanic Review* 16:1-32.

Rogers, Francis Millet. 1949. Insular Portuguese pronunciation: central and western Azores. *Hispanic Review* 17:47-70.

Sagey, Elizabeth. 1990. *The Representation of Features in Non-Linear Phonology: The Articulator Node Hierarchy*. Outstanding dissertations in linguistics. New York: Garland.

Samek-Lodovici, Vieri, and Prince, Alan. 1999. Optima. Ms. University of London, London, and Rutgers University, New Brunswick.

Sankoff, David, and Rand, D. 1988. *GoldVarb: a Variable Rule Application for the MacIntosh. Version 1.6.* Montréal: Centre de recherches mathématiques, Université de Montréal.

Santa Ana, Otto. 1991. *Phonetic Simplification Processes in the English of the Barrio: A Cross-Generational Sociolinguistic Study of the Chicanos of Los Angeles*. Ph.D. dissertation, University of Pennsylvania.

Santa Ana, Otto. 1996. Sonority and syllable structure in Chicano English. *Language Variation and Change* 8:63-89.

Schouten, M.E.H. 1982. T-deletie in de stad Utrecht: schoolkinderen en grootouders. *Forum der Letteren* 23:282-291.

Schouten, M.E.H. 1984. T-deletie in het zuiden van die provincie Utrecht. *Taal en Tongval* 36:162-173.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Selkirk, Elisabeth O. 1982a. The syllable. In *The Structure of Phonological Representations (Part II)*, eds. Harry van der Hulst and Norval Smith, 337-383. Dordrecht: Foris.

Selkirk, Elisabeth O. 1982b. *The Syntax of Words*. Cambridge: The MIT Press.

Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3:371-405.

Selkirk, Elisabeth O. 1987. On the nature of prosodic constituency. A comment on Beckman and Edwards. In *Papers in Laboratory Phonology*, eds. M. Beckman and J. Kingston, 179-200. Cambridge: Cambridge University Press.

Shulman, H.G., and Davison, T.C.B. 1977. Control properties of semantic coding in a lexical decision task. *Journal of Verbal Learning and Verbal Behavior* 16:91-98.

Silva, David James. 1988. The sociolinguistic variance of low vowels in Azorean Portuguese. In *Linguistic Change and Contact: Proceedings of NWAV-XVI*, eds. K. Ferrara, B. Brown, K. Walters and J. Bough, 336-344. Austin, TX: Department of Linguistics, University of Texas.

Silva, David James. 1991. Phonological variation in Korean: the case of the 'disappearing *w*'. *Language Variation and Change* 3:153-170.

Silva, David James. 1993. Chicano English evidence for the exponential hypothesis: a variable rule pervades lexical phonology. *Language Variation and Change* 4:275-288.

Silva, David James. 1997. The variable deletion of unstressed vowels in Faialense Portuguese. *Language Variation and Change* 9:295-308.

Silva, David James. 1998. Vowel lenition in São Miguel Portuguese. *Hispania* 81:166-178.

Smolensky, Paul. 1995. On the Internal Structure of the Constraint Component *Con* of UG. Talk presented at UCLA, April 7, 1995.

Smolensky, Paul. 1996. The Initial State and 'Richness of the Base' in Optimality Theory. Ms. Johns Hopkins University, Baltimore. [Technical Report JHU-CogSci-96-4.].

Spang-Hanssen, Henning. 1959. *Probability and Structural Classification in Language Description*. Copenhagen: Rosenkilde and Bagger.

Steriade, Donca. 1997. Phonetics in Phonology: The Case of Laryngeal Neutralization. Ms. UCLA.

Stevens, Kenneth N., and Blumstein, Sheila E. 1978. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 64:1358-1368.

Stevens, Kenneth N., and Keyser, Samuel Jay. 1989. Primary features and their enhancement in consonants. *Language* 65:81-106.

Stone, Gregory O., and Van Orden, Guy C. 1993. Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 19:744-774.

Sussman, Harvey M., McCaffrey, Helen A., and Matthews, Sandra A. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* 90:1309-1325.

Sussman, Harvey M., and Shore, Jadine. 1996. Locus equations as phonetic descriptors of consonantal place of articulation. *Perception and Psychophysics* 58:936-946.

Tesar, Bruce, and Smolensky, Paul. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29:229-268.

Tesar, Bruce, and Smolensky, Paul. 2000. *Learnability in Optimality Theory*. Cambridge: The MIT Press.

Tessier, Anne-Michelle. 2003. How and When to Project your Voice: Long-Distance Co-occurrence Restrictions, Feature Projections & Morphological Domains. Ms. University of Massachusetts, Amherst.

Tessier, Anne-Michelle. 2004. Root-restricted markedness and morpho-phonological domains. Paper presented at *Montreal-Ottawa-Toronto Phonology Workshop, February 2004, University of Ottawa*.

Thomas, Earl W. 1974. *A Grammar of Spoken Brazilian Portuguese*. Nashville: Vanderbilt University Press.

Tranel, Bernard. 1999. Optional schwa deletion: on syllable economy in French. In *Formal Perspectives on Romance Linguistics*, eds. J.-Marc Authier, Barbara E. Bullock and Lisa A. Reed, 271-288. Amsterdam: John Benjamins.

Treiman, Rebecca. 1983. The structure of spoken syllables: Evidence from novel word games. *Cognition* 15:49-74.

Treiman, Rebecca. 1986. The division between onsets and rimes in English syllables. *Journal of Memory and Language* 25:476-491.

Treiman, Rebecca, and Danis, Catalina. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language* 27:87-194.

Treiman, Rebecca, Gross, Jennifer, and Cwikiel-Glavin, Annemarie. 1992. The syllabification of /s/ clusters in English. *Journal of Phonetics* 20:383-402.

Treiman, Rebecca, Kessler, Brett, Knewasser, Stephanie, Tincoff, Ruth, and Bowman, Margo. 2000. English speakers' sensitivity to phonotactic patterns. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, eds. Michael B. Broe and Janet B. Pierrehumbert, 269-282. Cambridge: Cambridge University Press.

Twaddell, W.F. 1939. Combinations of consonants in stressed syllables in German. *Acta Linguistica* 1.

Twaddell, W.F. 1940. Combinations of consonants in stressed syllables in German. *Acta Linguistics* 2:31-50.

Ussishkin, Adam. 1999. The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output-output correspondence. *Phonology* 16:401-442.

Vennemann, Theo. 1972. Rule inversion. *Lingua* 29:209-242.

Vitevitch, Michael S., and Luce, Paul A. 1998. When words compete: levels of processing in spoken word recognition. *Psychological Science* 9:325-329.

Vitevitch, Michael S., and Luce, Paul A. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40:374-408.

Vitevitch, Michael S., Luce, Paul A., Charles-Luce, Jan, and Kemmerer, David. 1997. Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech* 40:47-62.

Walsh, Margaret A., and Diehl, Randy L. 1991. Formant transition duration and amplitude rise time as cues to the stop/glide distinction. *The Quarterly Journal of Experimental Psychology* 43A:603-620.

Wilson, Colin. 1999. Consonant cluster neutralization and targeted constraints. Ms. Baltimore, MD.

Wolfram, Walt. 1969. *A Sociolinguistic Description of Detroit Negro Speech*. Washington, D.C.: Center for Applied Linguistics.

Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: a loss of identity. *Linguistic Inquiry* 19:65-100.

Zubritskaya, Katya. 1997. Mechanism of sound change in Optimality Theory. *Language Variation and Change* 9:121-148.

Zuraw, Kie Ross. 2000. *Patterned Exceptions in Phonology*. Ph.D. dissertation, UCLA.