## Interacting (with) Morpheme Structure Constraints

Representational Solutions to Richness of the Base Problems in Optimality Theory

An der Philologischen Fakultät der Universität Leipzig<br>eingereichte

## DISSERTATION

zur Erlangung des akademischen Grades

Doctor philosophiae
(Dr. phil.)
von:
Sören Eggert Tebay

Gutachter_innen:
Prof. Dr. Jochen Trommer (Universität Leipzig)
Prof. Dr. Juliet Stanton (New York University)

Datum der Einreichung: 15.09.2021

Datum der Verleihung: 04.04.2022

## Acknowledgments

First and foremost, I want to thank my wife Delfina for her never-ending support and enduring love, and my son Willem for filling our live with joy.

I also want to thank the members of my dissertation committee for guiding me through each of the chapters and their areas of expertise with well-meaning assertiveness, while still giving me space for creative freedom. Special thanks to Juliet Stanton for becoming a member on such a short notice, to Jochen Trommer for accompanying me since the beginning of my career as a phonologist, to Eva Zimmerman for encouraging me to aim higher, and to Barbara Stiebels for bringing a typological perspective to the table.

I want to extend my gratitude to the past and present members of the faculty at InfL in Leipzig for creating such a wonderful atmosphere that really values collaboration and interaction, especially to Gereon Müller, Martin Salzmann, Phillip Weißer, Peter Staroverov, Martina Martinovič, and Nina Topintzi.

This dissertation would not have been possible without several people either hinting me at or providing me with linguistic data. I am grateful for fruitful discussion of data and analyses with Ezer Rasin, Laura Kalin, Sam Zukoff, Nik Rolle, Joana Serwaa Ampofo, Johanna Benz, and Aya Al-Ghanem.

I have also immensely profited from visiting scholars at our graduate program IGRA. I especially want to stress how thankful I am for the interaction with Laura Downing, Ricardo Bermudez-Otero, Peter Jurgec, Adam Albright, and Rachel Walker.

Last but not least, I owe a thank you to all my dear colleagues and fellow students at the IGRA graduate program and InfL for making this such a fun experience, especially preCorona. Special thanks my co-authors on various occassions Marie-Luise Popp, Christine Marquardt, Marie-Luise Schwarzer, and Masha Privizentseva, to my office mates Anna Bliß, Imke Driemel and Mike Frazier, and to Daniel Gleim for being the best phonology friend.

## Contents

Acknowledgments ..... i
List of Abbreviations ..... vii
1 MSCs, CoMDs, and CoURs ..... 1
1.1 OT, RotB and the Duplication Problem ..... 4
1.2 Morphoprosodic Hierarchical Structure ..... 6
1.3 Three Hypotheses on Phonological Domains ..... 14
1.4 Alternatives: CoMDs and CoURs in OT ..... 16
1.5 Overview of the Dissertation ..... 19
2 Typology of Monomorphemic Domains in Phonology ..... 23
2.1 Typological Hypotheses on CoMDs ..... 24
2.2 DoCoMD ..... 26
2.2.1 Entry Format ..... 27
2.2.2 Sampling ..... 29
2.2.3 Representativeness ..... 32
2.3 Exploratory Study ..... 33
2.3.1 Derived Phonological Properties ..... 33
2.3.2 Constraint Domains ..... 39
2.3.3 Naturalness of CoMDs ..... 43
2.3.4 Discussion of Exploratory Results ..... 48
2.4 Inferential Results ..... 48
2.4.1 Interaction of Properties ..... 49
2.4.2 Areal Distribution ..... 51
2.4.3 Discussion of Inferential Results ..... 53
2.5 General Discussion ..... 54
2.6 Conclusion ..... 59
3 The Domain Problem in Infixation ..... 61
3.1 Muna Labial Infix Dissimilation ..... 63
3.2 Hebrew Hitpa'el Metathesis ..... 66
3.3 Constraints on Non-Contiguous Domains ..... 68
3.4 Hierarchical Morphoprosodic Structure ..... 71
3.5 Overgeneration of Procedural Approaches ..... 79
3.5.1 SPE with Morpheme Structure Constraints ..... 80
3.5.2 Stratal Optimality Theory ..... 83
3.5.3 Harmonic Serialism ..... 86
3.5.4 Output-Output-Correspondence ..... 87
3.6 Undergeneration of Representational Solutions ..... 89
3.6.1 Boundary Markers ..... 89
3.6.2 Morphological Colors ..... 91
3.7 Crosslinguistic Data ..... 92
3.7.1 Austronesian Languages ..... 95
3.7.2 Afro-Asiatic Languages ..... 98
3.7.3 Other Language Families ..... 102
3.8 Conclusion ..... 104
4 Trigger Asymmetries in Vowel Harmony ..... 107
4.1 Headed Spans Theory ..... 110
4.2 Only High Triggers in Assamese ..... 113
4.3 No Short Mid Triggers in Päri ..... 119
4.4 Output Drivenness and Previous Approaches ..... 124
4.4.1 Constraints on Underlying Representations Ban Non-Triggers ..... 127
4.4.2 Root Strata Neutralize Non-Triggers ..... 129
4.4.3 Complex Faithfulness for Triggers Only ..... 130
4.4.4 Faithfulness in Headed Spans Theory ..... 134
4.5 Crosslinguistic Data ..... 136
4.6 Conclusion ..... 140
5 Richness of the Base Problems in Tonal Phonology ..... 143
5.1 Richness of the Base Problems ..... 144
5.2 Abstract Tonal Inventories in Chungli Ao ..... 150
5.2.1 Tonal Alternations in Bound Roots ..... 152
5.2.2 Recursive Prosodic Words as Tonal Domains ..... 155
5.2.3 (Non-)Footing in Prosodic Root Words ..... 158
5.2.4 Richness of the Base Revisited ..... 163
5.3 CoURs in Early Autosegmental Phonology ..... 165
5.3.1 Melodic Tonal Inventories in Mende ..... 165
5.3.2 Positional Tonal Inventories in Mee and Somali ..... 172
5.3.3 Universal Conventions in Autosegmental Phonology ..... 177
5.4 Crosslinguistic Data ..... 180
5.4.1 Tone in DoCoMD ..... 181
5.4.2 Category-specific Tonal Inventories ..... 183
5.5 Conclusion ..... 186
6 Discussion \& Conclusion ..... 187
6.1 Hierarchical Morphoprosodic Structure ..... 188
6.1.1 Richness of the Base and Output Drivenness ..... 188
6.1.2 Constraints on Derived Properties ..... 189
6.1.3 Diversity of Domains ..... 189
6.1.4 Interaction with Infixation and Vowel Harmony ..... 190
6.2 Stratal OT ..... 191
6.3 Output-Output Faithfulness ..... 192
6.4 Complex Faithfulness ..... 195
6.5 Constraints on Underlying Representations ..... 195
6.6 Conclusion ..... 196
Appendix: DoCoMD ..... 219

## List of Abbreviations

|  | Abbreviations |
| :---: | :---: |
| CoMD | constraint on monomorphemic domains |
| CON | constraint set in Optimality Theory |
| CoUR | constraints on underlying representations |
| DoCoMD | Database on constraints on monomorphemic domains |
| EVAL | evaluation component in Optimality Theory |
| GEN | generator in Optimality Theory |
| g -sample | genealogically controlled sample |
| HMS | Hierarchical Morphoprosodic Struture |
| HS | Harmonic Serialism |
| MSC | Morpheme Structure Constraint |
| MSR | Morpheme Structure Rule |
| OT | Optimality Theory |
| OT-OO | Optimality Theory with Output-Output Faithfulness |
| PR | phonological rule |
| RotB | Richness of the Base |
| SPE | Sound Pattern of English (Chomsky \& Halle 1968) |
| WALS | World Atlas of Language Structures |
| * | ungrammatical form derived by misapplication of processes |
| ** | unattested form due to excluded monomorphemic form |
| $\sigma$ | unattested, optimal candidate in an OT tableau |
| \% | attested, optimal candidate in an OT tableau |
| * | attested, non-optimal candidate in an OT tableau |
|  | Glosses |
| 1 | first person |
| 2 | second person |
| 3 | third person |
| ABESS | abessive case |
| ACC | accusative case |


| ACT | actor |
| :---: | :---: |
| ANTIP | antipassive voice |
| ATTR | attributive |
| DEF | definite |
| DEM | demonstrative |
| DETRANS | detransitivizer |
| EMPH | emphasis |
| ERG | ergative case |
| F | feminine |
| FOC | focus |
| FUT | future tense |
| INF | infinitive |
| INTR | intransitive |
| IPERF | imperfect tense |
| IRR | irrealis mood |
| LOC | locative case |
| M | masculine |
| NC | noun class |
| NMLZ | nominalizer |
| PERF | perfect tense |
| PL | plural number |
| PRS | present tense |
| PST | past tense |
| PTCP | participle |
| REA | realis mood |
| RECP | reciprocal voice |
| REFL | reflexive voice |
| SG | singular number |

## Optimality Theory Constraints

| AS | ADJACENTSpANS |
| :--- | :--- |
| CE | CrispEDGE |
| DEP | DEPENDENCY |
| F(AITH) | Faithfulness constraint |
| F(AITH)-IO | Input-Output Faithfulness constraint |
| F(AITH)-OO | Output-Output Faithfulness constraint |
| F(L) | FAITHFULNESS(Length) |
| FHS | FAITHFULNESSHEADSPAN |
| ID(ENT) | IDENTITY |


| M | Markedness constraint |
| :--- | :--- |
| MAx | Maximality |
| OCP | Obligatory ContourPrinciple |
| SH | *SpanHead |
| ŚH́ | SpánHÉÁd |
| SH-R | SpanHeadRight |
| SM | ShorteningMutation |
| \& | Local Constraint Conjunction |
| V | Constraint Disjunction |

## Phonological and Morphological Objects and Relations

| A/Aff | affix |
| :--- | :--- |
| A/ATR | advanced tongue root |
| BD | voiced obstruent cluster |
| BM | sequence of a labial obstruent and a bilabial nasal |
| C | consonant |
| Cor | coronal place of articulation |
| D | domination |
| DD | direct domination |
| Ft | foot |
| H | high tone |
| hi | high vowel |
| $\Im$ | identity relation |
| L | low tone |
| Lab | labial place of articulation |
| M | mid tone |
| N | noun |
| P/R | phonological material |
| Q | back or emphatic consonant |
| R/ | root |
| $\Re$ | correspondence relation |
| SM | shortening mutation |
| V | verb |
| $\alpha$ | a morphosyntactic constituent |
| $\mu$ | mora |
| $\pi / \rho$ | prosodic constituent |
| $\sigma$ | syllable |
| $\tau / v$ | tone |
| $\phi$ | cophonology |


| $\omega$ | prosodic word |
| :--- | :--- |
| $\omega$, NMIN | non-minimal prosodic word |
| $\sqrt{\omega}$ | prosodic root word |
| $([])$ | left or right edge of a domain |
| $\#$ | boundary marker |
| + | root boundary marker |
| . | syllable boundary |

## Chapter 1

## MSCs, CoMDs, and CoURs

Morpheme Structure Constraints (MSCs) have been used at least as early as Halle (1959). Their original purpose was to reduce redundancy in underlying representations. Underspecified values in underlying representations were filled in with predictable, redundant values by special rules that applied before any other morphological or phonological rules and applied in monomorphemic domains only. This usage continued in Sound Pattern of English (Chomsky \& Halle 1968). The notions of MSC was expanded to include negative constraints, conditional formulations and positive schemata by Stanley (1967). They have continued to play a major role in most phonological theories, barring Optimality Theory with its Richness of the Base principle, to be discussed in section 1.1. Their formal characterization has generally included three properties. ${ }^{1}$
(1) Defining properties of MSCs (cf. Kisseberth \& Kenstowicz 1977:136)
a. MSCs apply before any other morphological or phonological operations.
b. MSCs apply to monomorphemic forms.
c. MSCs are language-specific.

The practical application of MSCs generally fall into two groups. The first one is to explain phonological generalizations on monomorphemic domains. I will refer to these as constraints on monomorphemic domains (CoMDs). This challenge has been termed the Domain Problem (Kisseberth \& Kenstowicz 1977), cf. (2). The second one is to exclude certain inputs that would yield unattested outputs through application of the regular phonological grammar. This is often related to a context that is only identifiable

[^0]in underlying forms, but not in intermediate or surface representations. This challenge has been termed the Level Problem (Kisseberth \& Kenstowicz 1977), cf. (3). I will refer to this application of MSCs as constraints on underlying representations or CoURs. The main claim of this thesis is that CoURs are not needed for an empirically adequate theory of phonology, but CoMDs are necessary.
(2) Domain Problem (Kisseberth \& Kenstowicz 1977:149)

Some constraints seem to refer to the morpheme as it appears in the lexicon, other constraints refer to the structure that exists after words have been formed. These constraints refer either to the word or to the syllables that comprise the word.
(3) Level Problem (cf. Kisseberth \& Kenstowicz 1977:149)

At what level are constraints on phonological representations relevant? Is it the underlying structure? Or is it the phonetic structure?

Booij (2011:2052) provides an example from English for a constraint on monomorphemic domains. In English, morphemes can never end in a cluster of voiced obstruents, as shown in the examples (4-a,b). ${ }^{2}$ In morphologically complex words, however, such clusters are abundant, cf. (4-c,d). These systematic phonological generalizations in monomorphemic domains call for an explanation in any theory of phonology. One main claim of this thesis is that prosodic domains are sufficient domains for these generalizations.
(4) English constraints on voiced obstruent clusters (Booij 2011:2052)
a. **lıvd
b. ${ }^{* *} \mathrm{~d} \wedge \mathrm{bd}$
c. lnv-d
love-PST
d. d $\Lambda \mathrm{b}-\mathrm{d}$
dub-PST

The second application of Morpheme Structure Constraints has focused on phonological opacity, more specifically a certain pattern of counterfeeding interactions. In certain cases, it seems that a phonological generalization can be stated over the input of phonology, but not on its output, since the context is only present in the underlying form. A solution based on CoURs excludes inputs that would violate the generalization from the set of underlying forms, before any process could apply that would destroy the context.

[^1]Rasin (2016)'s analysis of assibilation in Finnish (Uralic, Finland) is a recent example of such an approach. Finnish assibilates the sequence /ti/ in derived environments to [si]. This includes a /t/before the past suffix -/i/ (cf. (5-a)) as well as before a word-final [i] that results from raising of an underlying /e/, cf. (5-b). As shown in example (5-c), the application of assibilation is blocked in underived environments, where /t/ precedes /i/ inside a morpheme.
(5) Finnish assibilation only in derived environments Rasin (2016:2)

```
a. [halusi]
    /halut/-/i/
    want-PST
    'wanted' (cf. [haluta] 'want.INF')
b. [vesi]
    /vete/
    'water' (cf. [vetenæ] 'water.ESSIVE.SG')
c. [tila]
    /tila/
    'room'
```

This context cannot be described in purely surface-oriented phonological terms. The process does not apply to any /ti/, but only to a specific subset of it. This can be easily described in terms of underlying representations. Any /ti/ sequence that is present in the underlying form is an exception to the assibilation rule. Rasin (2016) proposes that a CoUR demands underspecification of $/ \mathrm{t} /$ everywhere except before /i/ in the underlying representations. Full specification of /t/ before /i/ protects non-derived /ti/ sequences from assibilation. The assibilation rule applies to any underspecified /t/ that precedes an /i/ on the surface. ${ }^{3}$ I will propose that such patterns are derivable with richer representations in phonology. They do not require constraints on underlying representations.

In the following, I will continue to distinguish between constraints on underlying representations and constraints on monomorphemic domains. Both are taken to be language specific. Note that I will discuss a number of approaches, so that a variety of grammatical building blocks will fall under the two umbrella terms, just as people have used the terms before. Apart from OT constraints, I will also include, rules, schemata, conditions and

[^2]lexical strata with a specific ranking, as long as they adhere to the descriptions in (6) and/or (7).
(6) Defining Properties of CoURs
a. CoURs apply before any other morphological or phonological operations.
b. CoURs are language-specific.
(7) Defining Properties of CoMDs
a. CoMDs MSCs apply to monomorphemic forms.
b. CoMDs are language-specific.

One might wonder why we need a distinction between CoURs and CoMDs. This leads to an interesting question. Are CoURs enough to derive all CoMDs? I will argue that the answer is no. Certain phonological generalization refer to derived properties that are not present in the input, such as derived positions of infixes, derived segmental allophones and prosodic structure. This will be discussed in more detail in chapters 2 and 3.

As far as Optimality Theory is concerned, the reverse hypothesis is more attractive. Can all CoURs be explained using CoMDs? I will argue that they can, given sufficiently rich representations including Morphoprosodic Hierarchical Structure. Chapters 4 and 5 will give examples from tone and vowel harmony, using representational solutions to problems that have been described in terms of CoURs before. The remainder of this chapter will introduce the background on Optimality Theory and expand on the formal proposal Morphoprosodic Hierarchical Structure in sections 1.2 and 1.3. Alternative proposals to accommodate MSC-like effects in Optimality Theory will be discussed in section 1.4.

### 1.1 OT, RotB and the Duplication Problem

Optimality Theory (OT, Prince \& Smolensky 1993) is a parallel framework of phonology based on ranked violable constraints. The architecture of a phonological grammar is assumed to consist of three parts. A generator (GEN) takes a phonological input form and applies arbitrary changes to it, resulting in a set of potential output forms $\left\{\operatorname{Can}_{\mathrm{a}}, \operatorname{Can}_{\mathrm{b}}\right.$, $\ldots$...\}, known as candidates. This set of candidates is then evaluated by a special evaluation component (EVAL) against a constraint ranking (CON) $\left\{\mathrm{Con}_{1} \gg \mathrm{Con}_{2} \ldots\right\}$. A candidate surfaces as the output, known as the optimal candidate or winner, under certain conditions. Any candidate $\mathrm{Can}_{\mathrm{i}}$ is excluded if it violates a constraint $\mathrm{Con}_{\mathrm{x}}$, such that there in another candidate $\mathrm{Can}_{\mathrm{j}}$ that violates this constraint less severely and the other candidate $\mathrm{Can}_{\mathrm{j}}$ does not violate any higher ranked constraint $\mathrm{Con}_{\mathrm{y}}$ more severely than $\mathrm{Can}_{\mathrm{i}}$ does. Eventually, the only candidate that is not excluded becomes the actual output. From its inception,

Optimality Theory has been an output-oriented theory. This is especially evident from the Richness of the Base principle (RotB), which restricts the level of application of constraints in Optimality Theory, cf. (8). ${ }^{4}$
(8) Richness of the Base principle (Prince \& Smolensky 1993:191)

The set of inputs to the grammars of all languages is the same. (=There are no language-specific restrictions on the input. SET)

This principle is in direct conflict with MSCs and CoURs. In Optimality Theory, constraints can only apply to output forms and never to input forms. This idea is introduced as a solution to the Duplication Problem, which was identified in Kisseberth \& Kenstowicz (1977) and is given in (9).
(9) Duplication Problem (Kisseberth \& Kenstowicz 1977:136)

What appears to be a single phenomenon in some sense must be treated as two unrelated phenomena. (=As MSCs and phonological rules. SET)

In Optimality Theory there is no duplication of phonological information, since there is only one level for markedness constraints to apply due to RotB. All ranked markedness constraints apply to outputs, no such constraints apply to inputs. The Richness of the Base principle imposes a restriction on the set of constraints in OT. Constraints should only serve to exclude outputs or mappings from an input to an output. They can never exclude a certain input. Correspondence Theory (McCarthy \& Prince 1995) distinguishes between these two constraint purposes. Markedness constraints refer only to output structure and ban a certain marked structure. Faithfulness constraints on the other hand exclude certain mappings between corresponding structures from the input to the output, based on the changes between them.

Tesar (2013) further formalizes this notion as Output Drivenness and places an even tighter restriction on the set of possible constraints in Optimality Theory. More specifically, he excludes certain complex faithfulness constraints that could otherwise serve to exclude an input by mapping it to a unexpectedly marked or unexpectedly unfaithful candidate. In a positive way, it requires for OT that any mapping from an input I to an output O, implies the mapping of all inputs $\mathrm{J}, \mathrm{K}, \ldots$ to also map to the output O if $\mathrm{J}, \mathrm{K}, \ldots$ are properly more similar to the output O than I.

[^3]Output Drivenness (Tesar 2013:13)
A phonological map is output-driven if, for any mapping from an input to an output, any other input that has greater similarity to the output also maps to the same output.

I will assume that Output Drivenness, as defined in (10), is a sensible restriction on the constraint set of Optimality Theory. If we allow for mappings of certain inputs to any marked/unfaithful output in order to exclude them from our considerations, we are in effect restricting the relevant inputs for our analysis. In chapter 4, Output Drivenness will serve as the evaluation metric of approaches to asymmetric non-triggers in vowel harmony employing complex faithfulness constraints. The crosslinguistic typology of CoMDs in chapter 2 will show that constraints on monomorphemic domains are diverse and crucially make reference to derived properties, such as positional allophones and syllable structure. This can be explained by surface true markedness constraints, given the right representations.

Output Drivenness will also allow us to delineate more precisely, which kind of data could be problematic for an MSC-less theory like Optimality Theory with Richness of the Base. This is the case whenever an input needs to be mapped to an attested output even though there is another possible output with properly fewer faithfulness violations that can occur as the output for another input. Such an output should never become optimal because it loses on faithfulness and markedness grounds. A definition of such a generalized Richness of the Base problem is given in (11) and explained in more detail in chapter 5, where I argue that Richness of the Base problems can generally be solved by representationally distinguishing between outputs that otherwise look similar and introduce a markedness constraint that refers to these representations. In the next section, I will describe the representations needed.
(11) Richness of the Base problem

A potential input /A/ would map to an output [B] under the constraint ranking that has been established based on other input-output pairings and
a. empirically /A/ maps to $[\mathrm{C}]$ and
b. $[B]$ is independently well-formed and
c. the mapping $/ \mathrm{A} /[\mathrm{C}]$ is properly less faithful than the mapping $/ \mathrm{A} /[\mathrm{B}]$.

### 1.2 Morphoprosodic Hierarchical Structure

In this section, I will describe Morphoprosodic Hierarchical Structure, my proposal for prosodic structure of words and their relation to morphosyntactic features. These represen-
tation will provide the domains for constraints on monomorphemic domains. Additionally, they will serve to explain certain effects that have been attributed to constraints on underlying representations.

The representations in Morphoprosodic Hierarchical Structure are based on autosegmental representations of features and the prosodic-metrical hierarchy. Autosegmental Phonology, as introduced for tone by Leben (1973) and Goldsmith (1976), is based on the idea that phonological features are representationally autonomous. Their timing is not directly correlated with phonological segments. Instead, the relation is indirect, mediated via so-called association lines. These relate phonological segments and their features. Note that this relation does not have to be a one-to-one relation. In tone for instance, linking two tones to one segment will result in a contour tone (cf. (12-a)), whereas linking one tone to several segments will result in a span of segments that are all pronounced with a high tone, see (12-b).
(12) Autosegmental representation of tonal features
a. Contour Tone: Two tones linked to one segment

b. Tone Spans: One tone linked to two segments

H


Headed Spans Theory (McCarthy 2004) slightly extends autosegmental representations to adjust them for use with Optimality Theory. The main representational difference is that there are two kinds of relations between a segment and a feature. A segment can either be the head of a feature span or a non-head. Each span has to have exactly one head but a potentially unlimited number of non-heads. Simple spans are always headed by their only segment. Head status is assigned to output segments by GEN but is influenced by faithfulness constraints. This means that there are two possible interpretations of (12-b). Either it is headed by its first segment /á/ as in (13-a) or by its second segment /í/, as shown in (13-b). Notationally, heads are indicated by underlining and a bold association line. Additionally, spans are enclosed in brackets.

Headed Spans

(á í)

b.

In Chapter 4, I will argue that headed autosegmental spans allow us to derive asymmetric non-trigger effects in vowel harmony, a phenomenon that has previously been used as an argument in favor of constraints on underlying representations. The basic idea of the reanalysis is that certain vowels, the asymmetric non-triggers, can only occur if they are non-heads of a span. This excludes them from triggering vowel harmony and from appearing in isolation. Headed Spans will also play a role in the analysis of tones on bound verb roots in Chungli Ao in Chapter 5.

Whereas phonological features are autonomous, prosodic and metrical structure is organized hierarchically (Selkirk 1986; Nespor \& Vogel 2012). The inventory of potential constituents must include some statement that each constituent of level X has to dominate at least one constituent of level X +1 (cf. the Weak Layer Hypothesis in Itô \& Mester (2003)). Since the focus of this dissertation is on morpheme-level phonology, I will focus on word-internal structure. This is presented in (14). A prosodic word will always dominate at least one foot. The foot in turn, will always dominate at least one syllable and the syllable will always dominate at least one mora.
(14) Word internal prosodic hierarchy (Itô 1988; Inkelas 1990; Zec 1994)


Prosodic structure has been argued to include multiple prosodic words that can be in a dominance relation, also known as recursive prosodic words. ${ }^{5}$ Morphoprosodic Hierarchical

[^4]Structure incorporates and extends this idea. Each morpheme is potentially a prosodic word (van Oostendorp 1999; Raffelsiefen 2000) and therefore a possible domain for phonological computation. Additionally, minimal morphosyntactic information can be inherited by prosodic domains and therefore influence phonological computation.

First, consider the possible prosodic structures of a root with two suffixes into prosodic words. In the table in (15), root material is indicated by the square root symbol $\sqrt{ }$ and affix material is abbreviated as Aff. Each structure includes a root and two affixes.

In (15), the prosodic structures are labeled as (strictly) layered, coordinative and adjunctive, following Itô \& Mester (2021), cross-classified with binary or n-ary branching. In a layered structure, root and affix material is included in the same minimal prosodic word. I will leave this structure out of the following discussions, since it does not allow for systematic monomorphemic domains. Coordinative structures are different in that both the root and the affix are dominated by their own prosodic word, which in turn are dominated by another prosodic word. Coordinative structures are similar to layered structures in that the daughter constituents of the complex prosodic word are symmetric, i.e. of the same prosodic type. This is different in adjunctive structures. Here, only the root is included in its own prosodic word. Affixes are directly dominated by the next highest prosodic word. This means that the sisters dominated by this prosodic word are not of equal status and can be differentiated by their prosodic position.

Additionally, prosodic structures can be classified as being either binary branching or n-ary branching. Binary branching structures only allow each mother to dominate maximally two daughters. N -ary structure on the other hand include mothers with more than two children. This is crucial for adjunctive structures. In a binary adjunctive structure, each affix is directly dominated by its own prosodic word, whereas all affixes are dominated by the same prosodic word in an n-ary branching structure. I will show below that the former structure allows affixes to be treated as a domain for phonological processes.

[^5](15)

Possible Prosodic Structures


These prosodic constituents are important for a Morphoprosodic Hierarchical Structure because they will help to define domains for phonological markedness constraints. This will be based on two relations: domination and direct domination. Each of these relations defines a domain based on a prosodic constituent, in our case a prosodic word. Domination yields a domain that encompasses all material that is dominated by the prosodic word. Any intervening prosodic words can be ignored. It is thus different from direct domination. Direct domination defines a domain for a constraint that only ranges over material that is directly dominated by the constituent in questions without any intervening material of the same type. A definition is given in (16).
(16) Direct Domination (DD)

Phonological material P is directly dominated by a prosodic category $\pi$ if
a. $\pi$ dominates P and
b. there is no other prosodic category $\rho$ of the same type such that
(i) $\pi$ dominates $\rho$
(ii) $\rho$ dominates P

Three domains are crucial in the description of generalizations on monomorphemic domains. They will allow reference to affixes, roots and morphemes ${ }^{6}$ in general. The most general domain for a constraint is to apply to phonological material that is directly dominated by the same prosodic word, as in (17-a). This domain will allow reference to any monomorphemic domain. Other domains are more specific. If we define a non-minimal word as a prosodic word that dominates at least one other prosodic word, following Itô \& Mester (2021), we can define a new domain as phonological material being directly dominated by a nonminimal prosodic word. This will serve to identify affixes in general as a domain. Lastly, roots can be identified by a prosodic root word (cf. Inkelas 1990; Downing 1998, 2006). I will define this domain as phonological material (directly) dominated by the smallest prosodic word that includes the root. Note that the difference between direct domination and domination is not relevant for roots here, but see chapter 3 for an application.
(17) Three crucial domain specifications for constraints
a. ${ }^{*} \mathrm{XY}_{\omega, \mathrm{DD}}$ (=Prosodic Word, Direct Domination) Count one violation for every sequence XY that is directly dominated by the same prosodic word.
b. ${ }^{*} \mathrm{XY}_{\omega, \text { NMIN,DD }}(=$ Non-Minimal Prosodic Word, Direct Domination) Count one violation for every sequence XY that is directly dominated by the same non-minimal prosodic word (i.e. a prosodic word that dominates at least one other prosodic word).
c. $\mathrm{XX}_{\sqrt{\omega}, \mathrm{D} / \mathrm{DD}}$ (=Root prosodic word, (Direct) Domination) Count one violation for every sequence XY that is (directly) dominated by the smallest prosodic word that includes the root.

The actual practical application of these domains differs, based on the prosodic structure. The possible combinations are shown in the table in (18). Two crucial generalizations emerge. Fist, n-ary adjunctive structure is somewhat defective in that it does not allow reference to a general morpheme domain. Only roots are directly dominated by a prosodic words. Affixes share a prosodic word and can thus not be identified individually. The affix+affix domain is also picked out by the constraint domain $\omega$, NMIN,DD, as indicated by the dagger in both cases. ${ }^{7}$ The second important observation is that affixes can only be picked out by the $\omega$, NMIN, DD constraint domain. In all other structures, no single morpheme, root or affix, constitutes such a domain. Non-minimal words only dominate other prosodic words, except in the n-ary adjunctive structure, as mentioned above. ${ }^{8}$

[^6]Crucial constraint domains in Morphoprosodic Hierarchical Structure

|  | $\omega, \mathrm{DD}$ | $\omega, \mathrm{NMIN}, \mathrm{DD}$ | $\sqrt{\omega}, \mathrm{D} / \mathrm{DD}$ |
| :--- | :--- | :--- | :--- |
| binary, coordinative | morpheme | - | root |
| binary, adjunctive | morpheme | affix | root |
| n-ary, coordinative | morpheme | - | root |
| n-ary, adjunctive | root $\dagger$ | $-\dagger$ | root |

One might ask how morphological information, such as the root status is available for prosodic structure. I will argue that prosodic structure has to include minimal morphological information in order to account for category-specific effects in monomorphemic domains. In chapters 2 and 5 , I will show that several languages show such effects. This adds to the evidence provided by Smith (2011) for category-specific effects in phonology. While category-specific effects in general are a reasonable assumption (cf. e.g. Smith 2002, 2011), caution is advised. Lohmann (2017, 2018, 2020a,b) and Lohmann \& Conwell (2020) have recently shown that, as far as phonetic effects are concerned, lexical categories often only indirectly influence properties like duration, fundamental frequency and formant structure. Instead, these are affected by distributional properties. Verbs and nouns occur in different positions in sentences which influences the application of prosodic and intonational processes such as phrasal accentuation and phrase-final lengthening. Furthermore, verbs and nouns differ in type-frequency and token-frequency as well as predictability, which causes different degrees of reduction. This means that future research has to determine for each pattern if it can be explained by independent factors. Such a detailed examination lies beyond the scope of this thesis.

Category-specific effects will be modeled by Prosodic Category Inheritance. Morphological information on the lexical category, such as the noun-verb distinction as well as on root status are inherited by all morphoprosodic structure that includes roots bearing these features. The basic idea of Prosodic Category Inheritance is that prosodic constituents can be specified for categorical information. This category is inherited from the morphological root included in the prosodic constituent. If a verb root is included in a prosodic word, it will be a prosodic word associated with the categorical feature verb. If a noun root is included in a prosodic word, it will be a prosodic word associated with the categorical feature noun. Similar, root status is parasitically inherited. Only roots bear category features relevant for prosody. It follows that any prosodic word that is marked for a category feature includes at least one root. This simplifies the reference to roots in the definition of a prosodic root word. ${ }^{9}$

[^7]a. Prosodic Category Inheritance

The category of a prosodic constituent is inherited from a root that is included in the prosodic constituent.
b. Morphology $\longrightarrow$ Phonology

c. Root Inheritance Generalization

Any prosodic word that is marked for a lexical category contains at least one root.

After having established the representational proposal, we can turn back to the English example from the beginning. Recall that in English, morphemes cannot end in a cluster of voiced obstruents, but complex words can. This can be easily translated into a markedness constraint against clusters of voiced obstruents *BD at the end of some domain. This constraint needs to be relativized to the edge of a domain that includes all morphemes, i.e. for any phonological material directly dominated by the same prosodic word. The constraint definition is given in (20-a).
(20) Analysis of ban on English morpheme final voiced obstruent clusters
a. $\left.{ }^{*} \mathrm{BD}\right]_{\omega, \mathrm{DD}}$

Count one violation for any cluster of voiced obstruents that occurs at the right edge of a domain where all material is directly dominated by the same prosodic word.
b. Prosodic structure of $\mathrm{d} \Lambda \mathrm{bd}$ and $\mathrm{d} \Lambda \mathrm{b}-\mathrm{d}$


Morphoprosodic Hierarchical Structure allows us to distinguish between monomorphemic
main argument here. It remains an open empirical question how category-specific CoMDs interact with root compounding or category-changing affixes.
and polymorphemic forms. As shown in (20-b), a monomorphemic form violates the markedness constraint $\left.{ }^{*} \mathrm{BD}\right]_{\omega, \mathrm{DD}}$ because there is a voiced obstruent cluster at the right edge of the domain. Note that this problem does not occur in the structure of a polymorphemic form. Since the domain of the constraint only includes phonological material directly dominated by a prosodic word, the obstruent cluster is split up by a prosodic word boundary. Therefore the constraint is not violated and the polymorphemic form can become optimal in English whereas the monomorphemic form cannot. ${ }^{10}$ Analyses of these kind will be presented in chapters 3 and 5 .

### 1.3 Three Hypotheses on Phonological Domains

Even though we have established the representations in Morphoprosodic Hierarchical Structure, an important question remains. How do these structures come about at the morphology-phonology interface? I will not propose a new theory of prosodic mapping, since this would go beyond the scope of this thesis. Instead, I will discuss three hypothesis proposed for the morphology-phonology interface and sketch how an account based on established concepts can uphold all three hypotheses without losing its restrictiveness.

One of the most well-known hypotheses on the morphosyntax-phonology interface is the Indirect Reference Hypothesis, given in (21). It states that phonological rules can only make reference to prosodic structure and not directly to morphosyntactic structure. In Optimality Theory, that means that phonological markedness constraints can only refer to prosodic structure and not to morphosyntactic structure. It is important to note that some mapping constraints have to make reference to both morphosyntactic structure and prosodic structure in order to ensure the correct assignment of prosodic structure. The Indirect Reference Hypothesis can thus be maintained if two kinds of markedness constraints are distinguished. Phonological markedness constraints can only be specified for a prosodic domain. Prosodic markedness constraints can only refer to the mapping of prosodic structure and morphosyntactic structure. They cannot refer to other phonological material.

Indirect Reference Hypothesis (Inkelas 1990:10)
Phonological rules refer to only prosodic structure.

One possible theory of prosodic mapping is Match Theory (Selkirk 2011). Match constraints are responsible for mapping prosodic constituents to morphosyntactic constraints. They relate both edges of a morphosyntactic constituent to both edges of a prosodic constituent. Crucially, this mapping goes in both directions, from prosody to morphosyntax and

[^8]vice versa. Examples for morphemes (van Oostendorp 1999) and for roots are given in (22-a,b,c,d). I assume that detailed morphological affiliation is neither manipulable nor interpretable by phonology (van Oostendorp 2007). Phonology can only access lexical category and the root affix distinction for prosodic mapping.
(22) Some Prosodic Mapping constraints (Selkirk 2011:451,cf. van Oostendorp 2007)
a. Match(Morpheme,Prosodic Word)

Count one violation for each Morpheme whose left and right edges do not correspond to the left and right edges of a Prosodic Word in the output phonological representation.
b. Match(Prosodic Word, Morpheme)

Count one violation for each Prosodic Word whose left and right edges do not correspond to the left and right edges of a Morpheme in the output phonological representation.
c. Match(Root,Prosodic Word)

Count one violation for each Root whose left and right edges do not correspond to the left and right edges of a Prosodic Word in the output phonological representation.
d. Match(Prosodic Word, Root)

Count one violation for each Prosodic Word whose left and right edges do not correspond to the left and right edges of a Root in the output phonological representation.
e. Alternation(P,R)

Count one violation for any association line that is inserted between phonological material of the type P and R of the same morphological affiliation.

In addition to Match constraints, one further constraint type will be able to refer to morphosyntactic affiliation. These are Alternation constraints, which militate against insertion of association lines in tautomorphemic contexts. These constraints serve to derive certain kinds of derived environment effects, such as the Finnish data discussed above. Epenthetic material, such as the [+high] features on word-final raised vowels do not have a morphological affiliation. Similarly, [+high] features on affixes have a different morphological affiliation from roots. Linking these to a/t/ in the root therefore does not violate the alternation constraint. Non-derived /i/ inside roots on the other hand cannot be linked to a root-internal /t/ without violating the Alternation constraint.

Let us have a closer look at the second kind of constraints: phonological markedness constraints. How exactly can they make reference to prosodic domains? Jaker \& Kiparsky (2020) propose for independent reasons that phonological constraints should only refer to
domains in two ways: they can refer either to inclusion inside a domain or to the edge of a domain. This is termed the Domain Reference Hypothesis and provided in (23). Even though Jaker \& Kiparsky (2020) refer to morphological domains, the hypothesis can be translated to refer to prosodic domains. This means that phonological markedness constraints can either refer to the inclusion inside a domain or to the edge of a prosodic domain. The Domain Reference Hypothesis is thus in accordance with the present approach.
(23) Domain Reference Hypothesis (Jaker \& Kiparsky 2020:618)

The only morphological information to which phonological constraints may refer is morphological domains and their edges.

Accounting for MSC-like effects in phonology by prosodic or metrical units is not a new idea. Hooper (1972) attacked the idea of constraints on underlying representations by reintroducing the syllable into phonological theory. More recently, Gorman (2013) has questioned the presence of any static phonotactic constraints. He hypothesizes that there are no phonotactic constraints that cannot be explained by either phonological alternations (i.e. are dynamic) or by prosodic restrictions. Conversely, that means that all static phonotactic generalizations can be explained by reference to prosodic domains.
(24) No Static Phonotactics Hypothesis (Gorman 2013:11)

There are no static phonotactic constraints. All phonotactic constraints can be explained either by the phonological alternation system or by language specific prosodic inventory restrictions.

Morphoprosodic Hierarchical Structure is an extension of this hypothesis. There are static phonotactics (or phonological generalization on monomorphemic domains) and prosodic structure is used to explain all empirically attested generalizations.

### 1.4 Alternatives: CoMDs and CoURs in OT

Several alternative approaches have been proposed to deal with constraints on monomorphemic domains or constraints on underlying structures in OT. Here and in the following, I will mainly focus on three different strains of analysis: derivational versions of OT, exemplified by Stratal OT (Trommer 2011; Bermúdez-Otero 2012; Kiparsky 2015), outputoutput correspondence (Benua 1997; McCarthy 1998), and approaches based on complex faithfulness (Lubowicz 2002, 2003; Mahanta 2008).

Stratal OT can model MSC-like effects most easily. As already stated by Kiparsky (1985),
many generalizations on monomorphemic domains can also be stated over the stem. In Stratal OT this is implemented by having the specific markedness constraint high-ranked at the stem level, before some other affixes are added. Trommer (2011) adds a root stratum before the stem level to explain cases that cannot be analyzed as occurring at the stem level. Root stratum phonology evaluates candidates against a stratum-specific constraint ranking before any other morphological and phonological processes and applies only to monomorphemic domains. All morphemes are potentially evaluated on the root level. This allows for a simple analysis of English voiced obstruent clusters. A constraint against final voiced obstruent clusters *BD\# is high ranked at the root level but low ranked at any later level. This means that a monomorphemic voiced obstruent cluster is repaired at the root level, whereas a polymorphemic cluster is not yet present because the morphemes are evaluated separately at this level. At a later level, the creation of a cluster counterfeeds the repair mechanism for monomorphemic clusters at the root level.

Derivation table for a Stratal OT analysis of English voiced obstruent clusters

| Input | $\mathrm{d} \Lambda \mathrm{bd}$ | $\mathrm{d} \Lambda \mathrm{b}-\mathrm{d}$ |
| :--- | :--- | :--- |
| root level: ${ }^{*} \mathrm{BD} \# \gg$ FAITH | $\mathrm{d} \Lambda \mathrm{b}$ | $\mathrm{d} \Lambda \mathrm{b}, \mathrm{d}$ |
| later levels: FAITH $>*^{*} \mathrm{BD} \#$ | $\mathrm{~d} \Lambda \mathrm{~b}$ | $\mathrm{~d} \Lambda \mathrm{bd}$ |
| Output: | $\mathrm{d} \Lambda \mathrm{b}$ | $\mathrm{d} \Lambda \mathrm{bd}$ |

In chapters 3 and 4, I will argue that Stratal OT faces two problems. First, it overgenerates with respect to the empirically attested interactions between infixation and root domain constraints. Second, it is conceptually indistinguishable from Morpheme Structure Constraints as defined in the beginning of this chapter. A language-specific and level-specific constraint ranking applies only to monomorphemic domains and before any other morphological and phonological processes. It thus satisfies all defining properties of a Morpheme Structure Constraint. This makes it conceptually indistinguishable from other approaches with restrictions on the underlying representations. It severely restricts the input of any phonological computation proper and therefore violates the Richness of the Base principle.

A radically different approach to MSC-like effects is proposed in McCarthy (1998) couched in the framework of Output-Output-Correspondence (Benua 1997; OT-OO). The basic idea of Output-Output-Correspondence is to extend the notion of Input-Output correspondence from Correspondence Theory to a relation between a candidate and a morphologically related output, the so-called base. OO-faithfulness constraints can refer to this relation and enforce similarity between related forms even if not motivated by markedness constraints in the derived form. McCarthy (1998) makes use of this property by proposing that MSCs that take monomorphemic roots as their domain can be remodeled as OO-faithfulness
applying to isolated roots. This property of isolated roots can then be transferred to all derived forms due to OO-Faithfulness constraint. In other cases, the markedness constraint only applies to roots since its application is blocked by OO-Faithfulness constraints.

For an analysis of English voiced obstruents clusters, we make use of the latter type of analysis. We simply start with a constraint on word-final voiced obstruent clusters *BD\#. This constraint applies transparently to isolated roots excluding any root-final obstruent clusters because it is ranked above the Input-Output faithfulness. In the tableau in (26), this makes candidate ( $26-\mathrm{b}$ ) optimal, because candidate ( $26-\mathrm{a}$ ) violates the markedness constraint $* \mathrm{BD} \#$. In the optimal candidate, the constraint is repaired by rendering the cluster voiceless. The OO-faithfulness constraint does not become active, since isolated roots are not derived from a base.

Evaluation of an isolated root in OT-OO

| Input: <br> Base: <br> $\emptyset$ | FAIth-OO | *BD\# | Faith-IO |
| ---: | :---: | :---: | :---: |
| a. d $\Lambda \mathrm{bd}$ |  | $*!$ |  |
| b. d $\wedge \mathrm{pt}$ |  |  | $*$ |

In the morphologically complex derived form, the OO-faithfulness constraint becomes active and determines the optimal candidate. Changing the voicing value of the cluster, as in candidate (27-b), violates the OO-faithfulness constraint because it renders the root different from its base, the output form resulting from an input / $\mathrm{d} \wedge \mathrm{b} /$. The voicing value is different from its base, which incurs a fatal violation.

Evaluation of a derived form in OT-OO
$\left.\begin{array}{|rl||c|c|c|}\hline \begin{array}{r}\text { Input: } \\ \text { Base: }\end{array} & \mathrm{d} \wedge \mathrm{b} \text { b }\end{array}\right)$

In chapter 3, I will show that the Output-Output-Correspondence approach to MSC-like effects faces serious undergeneration problems. It is unable to derive attested patterns of interaction between infixation and root domain constraints because it only provides faithfulness for roots and not for affixes. More generally, as will be discussed in chapter 2, this is part of a severe problem. In OT-OO, only roots that can occur as isolated forms are potential domains for OO-faithfulness constraints. This means that neither bound roots nor affixes can serve as domains for constraints on monomorphemic domains. This is contrary to the empirical facts that emerge from the crosslinguistic study. Constraints
on monomorphemic domains need to be able to apply to affixes and roots independently.
Complex Faithfulness constraints have been used to exclude certain inputs by providing non-canonical mappings for them (Mahanta 2008). This allows for an emulation of certain effects of constraints on underlying representations. A partial analysis of the Finnish data can serve as illustration of the general idea. Recall that word-final raised [i] triggers assibilation, but underlying non-derived [i] does not. If complex faithfulness constraints are allowed to be specified for a condition in the input or the output, we can easily derive this pattern by the constraint given as a constraint conjunction in (28). This constraint is only violated if assibilation applies before a vowel that was high in the input.

$$
\begin{align*}
& \text { IdENT }[\text { sibilant }] \&_{\sigma} *[+ \text { high }]_{\text {Input }}  \tag{28}\\
& \text { Count one violation for a syllable that has a high vowel in the input and a } \\
& \text { consonant that changed its [sibilant] value from the input to the output. }
\end{align*}
$$

In chapter 5, I will argue that the analysis of asymmetric triggers in Assamese vowel harmony by Mahanta (2008), while empirically adequate, faces a conceptual problem. Complex faithfulness constraint systematically undermine the Output Drivenness of Optimality Theory. For a more detailed discussion see the specific chapter.

### 1.5 Overview of the Dissertation

The main goal of this dissertation is twofold. First, it provides a typological study of constraints on monomorphemic domains. Second, I will provide a theory that reconciles these empirical data with the principle of Richness of the Base in Optimality Theory. This will include empirical and conceptual arguments in favor of Morphoprosodic Hierarchical Structure and against competing approaches.

The first two chapters deal with the Domain Problem. They establish the typology of possible domains in phonology and deals with how Morphoprosodic Hierarchical Structure can be used for constraints on monomorphemic domains. Chapter 2 provides a crosslinguistic study of morphological domains in phonology. Some major findings include the following observations. Constraints on monomorphemic domains must be able refer to derived properties, such as allophones and syllable structure. In Desano for example, [d] and [r] are complementary allophones, with [d] occuring root-initally in [doa] 'sit' but not root-medially $* *[$ kede].

Domains are roots, affixes or any morpheme. In Muscat Arabic, back and emphatic consonants are banned from affixes, such that ${ }^{* *}$ - $\left[s^{\mathrm{s}} a \mathrm{ax}\right]$ is not a possible affix, but the same sequence occurs in the root [was ${ }^{〔} a x$ ] 'to be dirty'. In Dena'ina, simple roots cannot bear
an otherwise attested high tone, i.e. ${ }^{* *} / \int^{\text {j}} \mathrm{a}^{\mathrm{j}} /$ is not a possible simple root, but a possible complex word / ${ }^{\text {jáaj / meaning 'not go'. These data motivate Morphoprosodic Hierarchical }}$ Structure that allows independent reference to affixes, roots and morphemes. Additionally, a surface oriented approach to constraints on monomorphemic domains is motivated.

Chapter 3 provides a small scale typological study of the interaction between root domain constraints and infixation. The main generalization is that discontinuous roots created by infixation always remain domains for root structure constraints. An infix can violate root domain constraints, such as the Hebrew form /hi-s $\langle\mathrm{t}\rangle$ ater/ 'he hid himself', or be repaired to conform to it, as in Muna /p $\langle u m\rangle \mathrm{o}^{\mathrm{n}} \mathrm{ko} /[m o n g k o]$. Roots, on the other hand, show no such variation. No language exists where a root domain constraint only holds over non-infixed roots. This can be explained by a fixed ranking in Morphoprosodic Hierarchical Structure, but it poses a severe challenge to derivational approaches to constraints on monomorphemic domains.

The next two chapters provide reanalyses of data that have been used to motivate restrictions on underlying specifications. They deal with the Level Problem. I show that these data lend themselves to representational approaches. Chapter 4 deals with asymmetric non-triggers in vowel harmony in Assamese and Päri. In this pattern, certain vowels do not occur as triggers of vowel harmony and are banned in underived environments but can surface as the result of vowel harmony. In Assamese regressive [+ATR] harmony for example, non-triggering $-/ \varepsilon /$ ERG is an attested suffix, but a triggering counterpart -/e/ is unattested in suffixes. [e] can occur, however, if derived by vowel harmony, as in the form /pet/-/u/ [petu] 'pot bellied', compare unattested ${ }^{* *}$ [pete]. This has been used as an argument in favor of constraints on underlying representations, where these vowels are banned underlyingly but can occur if derived. I show that a simple representational analysis is possible based on headed spans. Asymmetric non-triggers can be banned from heading a feature span. This results in surface forms, where these vowels are banned from triggering vowel harmony or occurring in isolated forms. I also show that alternative approaches face conceptual problems and cannot adhere to the Richness of the Base principles and Output Drivenness.

Chapter 5 takes on the topic of Richness of the Base problems in tonal phonology. It includes a Morphoprosodic Hierarchical Structure analysis of the complex tonal patterns in bound verb roots in Chungli Ao, which have been used as an argument against Richness of the Base. In Chungli Ao, several Richness of the Base problems occur. One of them includes several melodies that can occur in derived forms but not in simple isolated roots. A low-mid melody, such as in the imperative form [cùmāy] cannot occur on a simple isolated root **[àcūm]. Additionally, melodic and positional inventories as commonly used in tonal analysis are shown to be reinterpretable in term of positional markedness
constraints referring to prosodic domains.
Chapter 6 summarizes the arguments in favor of Morphoprosodic Hierarchical Structure and against other approaches to CoMDs and CoURs, before concluding and giving an outlook of possible future research.

## Chapter 2

## Typology of Monomorphemic Domains in Phonology

This chapter provides a first exploratory study of phonological generalizations on monomorphemic domains. It is based on the database on constraints on monomorphemic domains (DoCoMD), which includes 229 constraints on monomorphemic domains from 140 languages. We have seen in chapter 1 that Morpheme Structure Constraints are usually defined by three properties. They apply before any other morphological or phonological operations, they apply to monomorphemic forms, and they are language-specific. An empirical typology of MSCs is inherently problematic because the decision if a phonological process applies before any other process in a grammar is a purely analytical decision that cannot be easily found in the data. Instead, I will focus on the monomorphemic domain and call these generalizations Constraints on monomorphemic domains (CoMD).
(1) Constraints on monomorphemic domains (=CoMD)

A phonological generalization that holds inside a monomorphemic domain, i.e. inside a root or an affx.

The example from English voiced obstruent clusters is repeated here and will serve as an illustrative example. In English, morphemes are restricted by a positional constraint on consonants. Inside a morpheme, a cluster of voiced obstruents cannot occur in final position, as shown by the impossible morphemes in (2-a,b). This constraint does not hold for whole words or syllables, since cluster of voiced obstruents are attested word-finally, as shown in (2-c,d).
(2) English constraints on voiced obstruent clusters (Booij 2011:2052)
a. ${ }^{* *} \operatorname{lnvd}$
b. ${ }^{* *} \mathrm{~d} \wedge \mathrm{bd}$
c. $\quad \operatorname{lnv}-\mathrm{d}$
love-PST
d. d $\Lambda \mathrm{b}-\mathrm{d}$
dub-PST

The remainder of this chapter is structured as follows. In section 2.1, I will present four hypotheses on COMDs that have been entertained in the literature and concern their substance, domain, and type. After presenting the database and explaining its format in section 2.2 , I will provide an exploratory study in section 2.3 , which already falsifies the hypotheses on substance and domain of CoMDs. Section 2.4 deals with the interaction of different properties of CoMDs and their areal distribution. This is followed by a general discussion in section 2.5 , where I show that the representational theory assumed in this dissertation can derive the crosslinguistic empirical data found in the exploratory study and contrast this with alternative theories of the phonology-morphology interface.

### 2.1 Typological Hypotheses on CoMDs

Different theories have often, either explicitly or implicitly, made restrictive hypotheses about possible constraints on monomorphemic domains (CoMDs). In this section, I will present four hypothesis on CoMDs that have been proposed in the literature. These hypotheses form the base of the typological study in this chapter.

Any theory that derives all constraints on monomorphemic domains as constraints on underlying representations (e.g. SPE, Chomsky \& Halle 1968), makes the strong prediction that a CoMD cannot refer to derived properties, as explicitly noted by McCarthy (1998). Derived properties include prosodic structure, such as moras, syllables as well as positional allophones. In subsection 2.3.1, I will show that Hypothesis 1 is empirically untenable. There are constraints on monomorphemic domains that need to make reference to derived prosodic structure as well as allophones.
(3) Hypothesis 1

Constraints on monomorphemic domains do not refer to derived phonological properties such as moras, syllables, feet, or positional allophones.

A second type of hypothesis concerns the possible monomorphemic domains that a CoMD can refer to. Surprisingly, contradictory ideas have been entertained in different theories.

Strictly cyclic theories, like Cophonology Theory (Inkelas 1998) or Lexical Phonology (Kiparsky 1982), exclude affix-specific phonological generalizations. The basic reasoning is that affixes enter phonological computation only accompanied by a base, i.e. they do not undergo any phonology in isolation and thus should not be subject to phonological generalizations that solely pertain to an affix domain. Similarly, OO-Correspondence Theory (Benua 1997; McCarthy 1998) holds that only potentially free forms should be subject to a CoMD. Since affixes are by definition bound, they should not be eligible as a domain for a CoMD.

## (4) Hypothesis 2a

Constraints on monomorphemic domains do not refer to proper affix domains.

A diametrically opposed prediction is explicitly made by the Root-Affix Faithfulness Metaconstraint (McCarthy \& Prince 1995). This metaconstraint on OT constraint rankings explicitly states that CoMDs can refer solely to an affix domain. Crucially, it also excludes any constraints referring only to a root domain. This is a consequence of ranking any faithfulness constraint that is indexed to affixes below its root-indexed version. In section 2.3.2, I will show that Hypothesis 2a and 2 b are both empirically untenable. CoMDs can refer to affixes, roots, or both.
(5) Hypothesis 2b

Constraints on monomorphemic domains do not refer to proper root domains.

A third hypothesis concerns the naturalness of CoMDs. Based on roots constraints in Lakohta (Siouan, USA), Albright (2004) argues that CoMDs are not always phonetically grounded, i.e. not all phonological generalizations on monomorphemic domains serve to ease articulation. Earlier proposals (Smith 2002) assumed that only sonorancy-related properties can be augmented in certain strong positions to ease parsing. Albright (2004) more generally assumes that constraints on roots can be morphologically grounded, i.e. they ease morpheme segmentation. These constraints tend to be constraints on possible sequences and morpheme edges. Under the assumption that consonant quality distinctions are perceptually more salient than vowel quality distinctions, this yields a statistical hypothesis on the clustering of CoMD properties.
(6) Hypothesis 3

Constraints on monomorphemic domains do not need to be phonetically grounded in articulation. In order to ease morphological parsing, root domain constraints tend to refer to sequences and/or positions of consonants.

In section 2.3.3, I will show that a modified version of Hypothesis 3 is confirmed by the genealogically controlled sample discussed in this study. Even though I am not aware of any existing theories on the geographic distribution, I will also discuss the geographic distribution of different properties of CoMDs and their influence on the validity of Hypothesis 3.

### 2.2 DoCoMD

DoCoMD is my Database on Constraints on monomorphemic domains. It includes 229 constraints on monomorphemic domains in 140 languages from 38 top-level language families from all over the world. In this section, I will describe the criteria used for an inclusion of a generalization into the database, describe the general format of DoCoMD and discuss the issues of sampling and representativeness.

A pattern is included in the database if it is an explicitly described, systematic phonological generalization that holds inside a monomorphemic domain. The pattern has to be explicitly described in a source that I have access to. I do not make inferences based on data provided in grammars or corpora. It also has to be explicitly restricted to a monomorphemic domain. In order to verify this, the following additional criteria were employed. First, if a constraint is only described for some monomorphemic domain, e.g. only in roots, I checked for any affixes violating this CoMD. This was especially easy for inventory constraints. A second test, applied especially in cases where a CoMD applies to all monomorphemic domains, was to look for whole words that would violate the constraint. This proved useful for different kinds of constraints. One way of putting this, is that a pattern is only included as a CoMD if it has exceptions outside of its domain. This excludes any alternative explanations of the constraint based on word or syllable level generalizations.

Systematicity is difficult to test in a quick and efficient manner. I did not perform any statistical analysis on phonological data of any particular language (cf. Gorman (2013) for such an approach). Instead, I looked at the quantity and quality of counterexamples to the generalization. If they were few in number or could be systematically derived by regular surface phonology, the pattern is still included.

### 2.2.1 Entry Format

An entry in DoCoMD consists of four different parts. These parts contain an index plus information about the language, the COMD, and the source. An example entry is given in (7).

Example entry for DoCoMD on English voiced obstruent clusters

| Index | eng1 |
| :--- | :--- |
| Language | English |
| Glottocode | stan1293 |
| Family | Indo-European |
| WALS-Genus | Germanic |
| Macro-Area | Eurasia |
| CoMD | Inside a morpheme, a cluster of voiced obstruents cannot <br> occur in final position. |
| Quotation | "English morphemes, for instance, never end in a cluster <br> of voiced obstruents" |
| Substance | Consonant <br> Type |
| Positional |  |
| Domain | Morpheme |
| Source | Booij2011:2052 |

Apart from an unique index of each entry, the first part of the entry pertains to the languages that the CoMD occurs in. The language name mentioned in the first source is provided in the field 'Language'. Information from the Glottolog (Hammarström et al. 2018) is also provided, namely the glottocode and the top-level family. Additionally, the WALS-Genus (Dryer \& Haspelmath 2013) might serve to ease genealogical sampling in future research based on this database. The geographic information is limited to the macro-area as used in the glottolog and defined by Hammarström \& Donohue (2014).

The actual CoMD is defined in the designated field. The formulation here is standardized and adheres to one of the positive or negative schemata in (8). (8-a) is a context-free formulation, whereas ( $8-b, c$ ) include context specifications. The difference between ( 8 -b) and ( $8-\mathrm{c}$ ) here is solely the scope of the context. Whereas (8-c) can describe generalizations that are obligatorily true in a certain context, (8-b) refers to generalizations that only hold in a certain context.
(8) Schemata for CoMD description
a. In DOMAIN, X must/cannot occur.
b. In DOMAIN, X must/cannot occur in CONTEXT.
c. In DOMAIN, in CONTEXT, X must occur.

Let us illustrate this difference with an example. In Naduhup (Naduhup; Brazil, Colombia) the three consonants $/ \mathrm{j}, \mathrm{g}, \varsigma /$ can only occur in morpheme-final position. This means that (9-a) with morpheme-final /¢̧/ is a possible root, but (9-b) with root-initial /¢̧/ is not. Crucially, other morpheme-final consonants are also allowed, such as /c/ in (9-c). The segments are thus restricted to a certain position, but the position can also be filled by other segments.
(9) Possible and impossible morphemes in Naduhup (Epps 2008:52)
a. /pǎç/
'stone'
b. **/çǎp/
c. /pác/
'mandube'

This is different in Ura (Oceanic; Vanuatu). In Ura, the verb-root-initial consonant position can only be filled by a restricted set of consonants $/ \mathrm{t}, \mathrm{s}, \mathrm{j}, \mathrm{w}, \mathrm{v} /$. Therefore, the verb in (10-a) with an initial $/ \mathrm{v} /$ is an attested verb root, but the verb root in (10-b) with an initial $/ \mathrm{k} /$ is not. $/ \mathrm{v} /$, however, is free to occur in a non-initial position, such as in a post-vocalic position on (10-c).
(10) Possible and impossible verb roots in Ura (Crowley 1999:149,151)
a. venim
'come'
b. ${ }^{* *}$ kenim
c. ivek
'defecate'

In the next field, a quoted description of the CoMD is given, inspired by the format in the Rara and Universals Archive (Plank \& Filimonova 2000). Note that the quotation is not always fully transparent due to differences in the use of certain terms, like morpheme, base, and stem. Each CoMD is also classified according to three properties: Substance, Type and Domain. The substance refers to kind of phonological objects targeted by the CoMD.

These can be Consonant, Vowel, Consonant+Vowel, Tone, Mora and Syllable. As for types of CoMDs, I distinguish inventory constraints, sequence constraints, positional constraints and minimality or maximality constraints. Inventory constraints do not come with any contextual specification and simply ban the occurence of some phonological object in a certain domain. Sequential constraints on the other hand refer to the cooccurence of certain phonological objects either in a direct precedence relationship or more generally in a precedence relationship. Positional constraints restrict a certain object or a sequence of objcects to a certain position, relative to its domain. Such constraints often refer to intial or final positions in their domain.

Additionally, the kind of domain is surveyed. This field cross-classifies the lexical category noun (N) vs. verb (V) with the root-affix distinction. In effect, a CoMD can either apply to roots, to affixes, or to any monomorphemic domain irrespective of its status as a root or affix. An additional distinction between root and root proper is needed because many sources discuss constraints on roots without mentioning their status for affixes. In other cases, the CoMD cannot apply to affixes for independent reasons. The value root proper on the other hand is assigned when the constraint explicitly and concretely only applies to roots and not to affixes.

| Possible values for CoMD properties |  |
| :--- | :--- |
| Field | Possible values |
| Substance | Consonant, Vowel, Consonant+Vowel, Tone, Mora, Syllable |
| Type | Inventory, Sequence, Positional, MinimalityMaximality |
| Domain | root, root proper, affix, morpheme |
|  | N root, N root proper, N affix, N morpheme |
|  | V root, V root proper, V affix, V morpheme |

Finally, a list of sources with page numbers is given. In most cases, the list only includes a single source. The full database is found in the appendix of this dissertation.

### 2.2.2 Sampling

The starting point for the database were the distributional generalizations present in the P-base database (Mielke 2008). The description of these distributions was searched for the following terms: root, stem, morpheme, and affix. The resulting set of entries was double checked with the given source if available and included if it satisfied the inclusion criteria. This resulted in 44 entries.

The database was then expanded in order to obtain a variety sample, i.e. languages were supposed to be "represented in proportion to family and area sizes" (Bickel 2008:222).

Such a sample is especially valid for exploratory studies on topics where crosslinguistic data are missing so far (cf. Bakker 2010). This is true for the crosslinguistic distribution of CoMDs. The whole database will serve as the basis for the exploratory study in section 2.3. However, when more data on the interaction of CoMDs with infixation and vowel harmony (cf. chapter 3 and 4) were collected, a possible bias towards Austronesian and Atlantic-Congo languages was introduced and the database in whole is thus much more of a convenience sample. Note that this is less problematic for an explorative study where over-attestedness is not an issue.

In order to control for a genealogical bias, the post-hoc procedure developed by Bickel (2008) was employed to extract a genealogically controlled sample, a so-called g-sample, from the database. This process involved several steps. First, for every top-level family, the number of daughter languages in the database was determined. If a top-level language only had one daughter language in DoCoMD, it was directly included in the $g$-sample. If there were several daughter languages, these daughter languages were compared. If the sets of CoMDs associated with these languages had identical values for the properties of substance, type and domain, only one of these languages was included. ${ }^{1}$ If the daughter languages were not identical, the next-lower sub-families were subjected to the same steps. This method excluded several languages with identical CoMDs from subfamilies of Austronesian, such as Greater Central Philippine and North Borneo, from subfamilies of Atlantic-Congo, such as Ubangi, Southern Bantu-Makua, Oti North Guang and Great Lakes Bantu as well as from smaller families that were overrepresented before, such as the Central Sudanic languages. The resulting g-sample includes 200 entries from 111 languages.

| Macro-areas in DoCoMD and g-sample |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
|  | DoCoMD | g-sample |  | Glottolog |  |  |
|  | $\#$ | $\%$ | $\#$ | $\%$ | $\#$ | $\%$ |
| Africa | 63 | 45 | 38 | 34 | 2348 | 28 |
| Australia | 5 | 4 | 5 | 5 | 386 | 5 |
| Eurasia | 28 | 20 | 28 | 25 | 1971 | 23 |
| North America | 9 | 6 | 9 | 8 | 788 | 9 |
| Papuanesia | 28 | 20 | 24 | 22 | 2209 | 26 |
| South America | 7 | 5 | 7 | 6 | 711 | 8 |
| Total | 140 | 100 | 111 | 100 |  | 100 |

The representation of different macro-areas is given in (12). The number are compared

[^9]to the proportion of languages from this macro-area in the Glottolog on April 29th, 2021, assuming a total of 8516 languages. In the following, all statements of over- and underrepresentation will be relative to the Glottolog numbers. In the database, Africa is overrepresented with 63 of $140(63 \%)$ languages in the database. As mentioned above, this is mainly due to the inclusion of several entries that interact with vowel harmony. Even though 25 languages of Africa have been excluded for the g -sample, these are still overrepresented in the g -sample. In the database, all other macro-areas are approximately equally underrepresented according to their relative number of languages in the Glottolog. In the g-sample, Eurasia, Papuanesia and South America are slightly less represented than what would be expected from their proportional share of the languages of the world.

| Top-level families in DoCoMD and g-sample |  |  |  |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | :---: | :---: | :---: |
|  | DoCoMD | g-sample |  |  |  |  |  | Glottolog |  |
|  | $\#$ | $\%$ | $\#$ | $\%$ | $\#$ | $\%$ |  |  |  |
| Atlantic-Congo | 38 | 27 | 20 | 18 | 1436 | 17 |  |  |  |
| Austronesian | 24 | 17 | 20 | 18 | 1276 | 15 |  |  |  |
| Indo-European | 9 | 6 | 9 | 8 | 588 | 7 |  |  |  |
| Central Sudanic | 9 | 6 | 3 | 3 | 64 | 1 |  |  |  |
| Nilotic | 7 | 5 | 7 | 6 | 55 | 1 |  |  |  |
| Sino-Tibetan | 6 | 4 | 6 | 5 | 496 | 6 |  |  |  |
| Afro-Asiatic | 6 | 4 | 6 | 5 | 375 | 4 |  |  |  |
| Nuclear Trans New Guinea | 4 | 3 | 4 | 4 | 316 | 4 |  |  |  |
| Pama-Nyungan | 3 | 2 | 3 | 3 | 248 | 3 |  |  |  |
| Uralic | 3 | 2 | 3 | 3 | 48 | 1 |  |  |  |
| Otomanguean | 2 | 1 | 2 | 2 | 1818 | 2 |  |  |  |
| Surmic | 2 | 1 | 2 | 2 | 11 | 0 |  |  |  |
| Koman | 2 | 1 | 1 | 1 | 5 | 0 |  |  |  |
| Other | 25 | 18 | 25 | 23 | 873 | 10 |  |  |  |
| Total | 140 | 100 | 111 | 100 |  |  |  |  |  |

A list of top-level language families present in DoCoMD and the extracted g-sample is provided in (13). ${ }^{2}$ In total, 38 top-level families are included. Even though the AtlanticCongo family is the largest top-level family in the Glottolog, it is vastly overrepresented in the database. Sino-Tibetan languages on the hand are slightly underrepresented, again compared to their relative number of daughter languages. This is mostly due to a linguistic

[^10]fact about the languages found in certain subfamilies of Sibo-Tibetan, especially Sinitic languages. In these languages morphemes are often coextensive with the word or the syllable and therefore alternative explanations for phonological constraints supposedly holding for monomorphemic domains cannot be easily excluded. The same applies to other South East Asian language families not included in this study, e.g. Hmong-Mien and Tai-Kadai. Several smaller families are overrepresented because there are more data available and the languages feature CoMDs that interact with vowel harmony. These are Central Sudanic, Koman, Nilotic, Surmic, and Uralic.

In the g-sample, several Atlantic-Congo and Austronesian languages have been excluded. Both language families are now represented roughly proportionally. Central Sudanic, Nilotic, Surmic, and Uralic are still overrepresented due to their internal diversity with regards to CoMDs.

### 2.2.3 Representativeness

DoCoMD can only be considered representative for the languages of the world under certain conditions and for certain purposes. This has several reasons. First and foremost, constraints on monomorphemic domains are not equally well described for different languages. A typical phonotactics section in a descriptive grammar consists of subsections on syllables and phonological words. If description is carried out adhering to such a scheme, phonological constraints on monomorphemic domains in a given language might be missed. Therefore, absence of evidence for a certain CoMD cannot be used as evidence for the absence of such a CoMD.

On the other hand, the presence of a certain CoMD can be stated with considerable certainty only depending on the quality and accuracy of the description provided in the sources used. This means that the explorative study is likely to uncover typologically valid results, whereas the interaction of CoMD properties as surveyed should only serve as a means of formulating more succinct, testable hypotheses.

Since no statistical tests were used in the study to test the significance of CoMDs in a single language, no statements on gradient CoMDs (cf. e.g. Frisch et al. 2004) were included. ${ }^{3}$ Similarly, no claims about the psycholinguistic reality of CoMDs can be made. It has been known at least since Zimmer (1969) that some CoMDs found in descriptive work cannot be replicated in linguistic experiments. Gorman (2013) shows that this true even for some statistically significant generalizations. DoCoMD only includes categorical statements since sources usually were not detailed enough to make more precise statements.

[^11]
### 2.3 Exploratory Study

In the following section, I will describe an exploratory study based on DoCoMD. In the course of this, I will show that Hypothesis 1, Hypothesis 2a, and Hypothesis 2b can be easily falsified. The study will arrive at three conclusions. CoMDs on derived phonological properties exist. Both affixes and roots can serve as domains for CoMDs. The motivation for CoMDs can be either found in articulatory phonetics or in morphological parsing. I will also describe some concrete data from diverse languages that support these conclusions.

### 2.3.1 Derived Phonological Properties

In order to survey derived phonological properties, I will discuss the phonological substance and the constraint types found in the languages in DoCoMD. The frequencies of the different values for constraint substance are given in (14).

| Constraint Substance in DoCoMD |  |
| :--- | ---: |
| consonant | 101 |
| vowel | 88 |
| consonant+vowel | 17 |
| syllable | 10 |
| tone | 9 |
| mora | 4 |
| total | 229 |

CoMDs on consonants are the most common, followed by constraints on vowels. Overall, constraints on segments make up the majority of the database ( $\mathrm{n}=206$ ). However, constraints on non-segmental material exist ( $\mathrm{n}=23$ ). These can refer to syllables, moras, and tone.

Take for example roots in Ixpantepec Nieves Mixtec (Otomanguean; Mexico). These are subject to a constraint that requires them to be exactly bimoraic. This means that they can consist of a single heavy syllable or two light syllables but neither of a single light syllable nor of two heavy syllables or more. A heavy syllable is one that includes a long vowel or a coda consonant. In (15), some attested and impossible roots are given. ${ }^{4}$ (15-a,b) show roots with two light syllables, whereas ( $15-\mathrm{c}, \mathrm{d}$ ) consist only of one heavy syllable. The roots in (15-e,f,g) are not attested because they consist of three light syllables, a light and a heavy syllable, or four light syllables, respectively.

[^12](15) Attested and impossible roots in Ixpantepec Nieves Mixtec (Carroll 2015:56)
a. [?uั̀nï]
/ùnì/
'three'
b. [niั̀nò̀]
/nìnu/
'up'
c. $[\overline{\overline{11}}]$
/iil
'one'
d. [nũ̃ $\overline{\tilde{u}}]$
/nuu/ 'town'
e. ${ }^{* *}\left[k^{w}\right.$ ànã̄nā]
/kwànana/
f. $* *\left[\mathrm{t}^{\mathrm{j}} \mathrm{k}^{\mathrm{w}} \mathrm{miñ}_{1 \mathrm{i}}\right]$
$/ \mathrm{t}^{\mathrm{j}} \mathrm{ik}^{\mathrm{w}}{ }^{\mathrm{I} \mathrm{i}} /$
g. ${ }^{* *}[$ ūhū̀'válí]
/uxuválî/

Note that this constraint does not hold for words. Words can be longer than two light syllables, as exemplified by the data in (16). Affixes can also be shorter than two moras, i.e. consist of a single light syllable. The constraint therefore holds purely for roots.
(16) No size constraint on complex words in Ixpantepec Nieves Mixtec
(Carroll 2015:40,46,128)
a. [ $\mathrm{k}^{\mathrm{w}} \mathrm{a}^{2}$ ã̃ā]
/kwà-nana/
MOTION-go_up
'go up'
b. $\left[\mathrm{t}^{\mathrm{j}} \mathrm{i}^{\mathrm{w}}{ }^{\mathrm{IIII}}\right]$
$/ \mathrm{t}^{\mathrm{j}} \mathrm{i}-\mathrm{k}^{\mathrm{w}}{ }_{\mathrm{i}}^{\mathrm{i}} / \mathrm{L}$
NOUN.CLASS-mosquito
'mosquito'
c. [?ūhū'válí]
/uxu-válî/
deer-little.PL
'fawns'

In Ejagham (Atlantic-Congo; Cameroon, Nigeria) on the other hand, the number of syllables occuring in a morpheme is restricted. Morphemes can be either disyllabic or monosyllabic. This means that morphemes consisting of less than one syllable, i.e. a sole non-syllabic consonant, are banned as well as morphemes consisting of three syllables or more. Note that words can become longer than two syllables. Some existing and impossible roots in Ejagham are given in (17) and some morphologically derived trisyllabic words in (18).
(17) Possible and impossible morphemes in Ejagham (Watters 1981:33,34)
a. /à/-

3SG-
'(s)he/it'
b. $-/ t \hat{a} /$
-sting
'sting'
c. -/pínì/
-tuble
'tumble'
d. /í/-

NOUN.CLASS-
e. -/fúdí/
-cane.rope
'cane rope'
f. ${ }^{* *}$ b-
g. **apini
h. ${ }^{* *}$ ifudi
(18) No size constraint on words in Ejagham (Watters 1981:33,34)
a. [àphínì]
/à/-/pínì/
3sG-tumble
'he tumbled'
b. [ífúrí]
/í/-/fúdí/
NOUN.CLASS-cane.rope
'cane rope'

These two were examples of constraints on the size of monomorphemic domains. However,
other constraint types also make reference to derived prosodic structure. In order to discuss these, let us first look at the frequencies of different constraint types, as shown in (19).

| Constraint Types in DoCoMD |  |
| :--- | ---: |
| sequence | 73 |
| inventory | 72 |
| positional | 71 |
| minimality/maximality | 13 |
| total | 229 |

The most frequent constraint types are sequence constraints, inventory constraints and positional constraints, with similar frequencies. Minimality/maximality are considerably less frequent in DoCoMD. However, minimality/maximality constraints are not the only constraints that can refer to derived phonological properties.

In several languages, the inventory of possible syllables in monomorphemic domains in restricted. Such a constraint exists in Lakhota (Siouan, USA) against closed syllables in roots, in Yeri (Nuclear Toricelli, Papua New Guinea) against superheavy syllables in roots, and in Milang (Sino-Tibetan, India) against superheavy syllables in any morpheme. Superheavy syllables are those that include a long vowel as well as a coda consonant. Examples from Milang are given in (20) and (21). The forms in (20-a,b) are possible morphemes because they include a single light syllable. Similarly, a closed heavy syllable with a coda consonant in (20-c) and an open syllable with a long vowel in (20-d) are allowed. In contrast, the superheavy syllables in (20-e,f) with both a long vowel and a coda consonant are not attested as morphemes. Nevertheless, the same forms show up as morphologically complex words in (21-a,b), again with a long vowel and a coda consonant.
(20) Possible and impossible morphemes in Milang (Modi 2017:94)
a. $u$

DEF
b. ka-
'look'
c. man-
'roast'
d. bjaa

HESTERNAL
e. ${ }^{* *}$ dzaam
f. ${ }^{* *}$ joon
(21) Superheavy syllables in complex words in Milang (Modi 2017:95)
a. dzaam
dzaa-um
3.PL-ACC
'them'
b. joon
joo-n
HYPERDISTAL.SAME_LEVEL-DEM.FOC
'It's up there.'

Positional constraints also frequently refer to syllable structure. ${ }^{5}$ This can be illustrated with data from Semelai (Austroasiatic, Malaysia). In Semelai, the coda consonants of penultimate syllables in roots are severely restricted. Only nasals and liquids are allowed, as shown in (22-a,b,c) vs. (22-d,e).
(22) Possible and impossible roots in Semelai (Kruspe 2004:42)
a. [kəm.pən]
/km.pən/
'wife'
b. [pəl.to?]
/pl.tor/
'explosion'
c. [kər.te?]
/kr.te?/
'personality'
d. ** racbac
e. ${ }^{* *}$ kәрляр

Note that this is not a general restriction on syllable structure, since final syllables can end in a larger array of syllables, as shown in (22-b,c). It is also not a restriction on words, since a morphological operation of infixing reduplication can create coda consonants of the penultimate syllable that violate this constraint. The complex words given in (23) are examples of such violations.

[^13]（23）No coda constraint on complex words in Semelai（Kruspe 2004：73，74）
\[

$$
\begin{array}{ll}
\text { a. } & {[\text { rac. bac] }} \\
& \text { /ra }\langle\mathrm{c}\rangle \text { bac/ } \\
& \langle\mathrm{INTR}\rangle \text { scratch } \\
& \text { 'to be scratching' }
\end{array}
$$
\]

b．［kәр．nєp］
／k〈p〉nep／
〈INTR〉blink
＇to be blinking＇

Finally，some constraints refer to allophones，i．e．to sounds that are in complementary distribution．An illustrative example is found in Desano（Tucanoan；Brazil，Colombia）．In Desano，［d］cannot occur in root－medial position．In this context，its positional allophone ［r］occurs．In（24－a，b，c）［d］occurs in root－initial position，whereas in（24－c，d）［r］occurs in root－medial position．Examples（24－e，f）show that［d］cannot occur root－medially，since these forms are unattested．
（24）Possible and impossible roots in Desano（Silva 2012：42）
a．［doa］
／doa／
＇sit＇
b．［de ${ }^{\mathrm{h}} \mathrm{ko}$ ］
／deko／
＇water＇
c．［dare］
／dade／
＇ipixuna fish＇
d．［kere］
／kede／
‘jacu bird＇
e．${ }^{* *}$ weadigt
f．＊＊moaduagu

This is not a general restriction on words，since roots can occur in a non－initial position in words．In these cases，［d］still surfaces in root－intial position．In the non－initial roots in （25），［d］still surfaces root－initially．
(25) Word-medial [d] in Desano (Silva 2012:43)
a. [weadigu]
/wea-di-gu/
clay-meat-3SG.m
'statue'
b. [mõãduagu]
moa-dua-gu
build-want-3SG.m
'(he) wants to build'

The data presented above show that CoMDs make reference to syllables, moras and contextual allophones. These counterexamples thus falsify the Hypothesis 1, which states that such CoMDs should not exist. This is a problem for theories that assume that any constraint on monomorphemic domains should be derived by constraints on underlying representations, where predictable prosodic and allophonic information is not present. Any theory of phonological constraints on monomorphemic domains must include some way to explain constraints on derived phonological properties.

### 2.3.2 Constraint Domains

Hypothesis 2 a and 2 b concern the availability of different kinds of domains for phonological generalizations. In this subsection, I will show that both affixes and roots are available, thereby falsifying both hypotheses. In order to do this, we need to examine the distribution of different domains in the dabase. The counts of different domains in DoCoMDs are given in (26).

Constraint Domains in DoCoMD

|  | root | root proper | affix | morpheme | total |
| :--- | ---: | ---: | ---: | ---: | ---: |
| noun | 4 | 1 | 1 | 0 | 6 |
| verb | 8 | 1 | 0 | 0 | 9 |
| both | 85 | 19 | 11 | 99 | 214 |
| total | 97 | 21 | 12 | 99 | 229 |

There is an obvious generalization that can be drawn from this table. The most common domains are roots and morphemes. Together, they make up the majority of the database ( $\mathrm{n}=196$ ). This means that a large amount of constraints refer to both roots and affixes or refer to roots without any information on their application to affixes.

In order to test the hypothesis, the existence of constraints on roots proper and affixes are
crucial. One simple example for a constraint on proper roots is from Lakhota. In Lakhota only open syllables are allowed in roots, cf. (27). In affixes and affixed words, closed syllables can occur, as shown in (28). These closed syllables come about (among other processes) by apocope of word-final /i/ included in affixes ${ }^{6}$ in (28-a,b) or by attaching the short form $[k J]$ of the emphatic suffix $/ \mathrm{kjto} /$ in (28-c).
(27) Possible and impossible roots in Lakhota (Albright 2004:2)
a. $\mathrm{tk}^{\mathrm{h}} \mathrm{a}$
'but'
b. gnæ
'cheat, fool'
c. ${ }^{* *}$ [kat $]$
d. ${ }^{* *}[\operatorname{tax}]$
e. ${ }^{* *}[\operatorname{man}]$
(28) Closed syllables in complex words in Lakhota (Albright 2004:4)
a. [ju.hap]
/juha/-/pi/
have-PL
b. [ju.hak]
/juumha/-/ki/
have-DEF
c. [nijekf]
/nije/-/kj/
you-EMPH
'it's up to you'

Root domain constraints can also refer to suprasegmental material, such as tones. In Dena'ina (Athabaskan-Eyak-Tlingit, USA) roots cannot bear a high tone. This tone is only attested in the negative form of verbs and in certain directionals. As shown in (29), low tones and mid tones, but not high tones, are attested in simple roots. This contrasts with the negative forms in (30), where a high tone is obligatory. Note that the morphological make-up off the negative forms is not transparent. Negation is additionally marked by deaspiration of final $/ \mathrm{t}^{\mathrm{h}} / \mathrm{in} / \mathrm{kót} /$ and voicing of $/ \mathrm{h} /$ to $/ \mathrm{j} /$ (Lovick 2020:132). Note that negative forms also appear with additional segmental negation markers (Lovick 2020:459,634).

[^14](29) Possible and impossible roots in Dena'ina (Lovick 2020:76)
a. $/ \int^{j} \bar{a} h /$
'go'
b. /kòt ${ }^{\mathrm{h}} /$
'punch'
c. ${ }^{* *} / \mathrm{S}^{\mathrm{j}} \mathrm{a}^{\mathrm{h}} /$
d. ${ }^{* *} /$ kót $^{\text {h }} /$
(30) High tone in complex negative forms in Dena'ina (Lovick 2020:76)
a. / ${ }^{\text {jáj }}$ /
'not go'
b. /kót/
'not punch'

For any constraint on affixes, it is crucial to establish that they cannot be derived by more general constraints. A well known example is an inventory constraint on consonants in suffixes in several Arabic varieties. In Muscat Arabic (Afro-Asiatic, Oman) for example, emphatic and back consonants $/ \delta^{£}, \mathrm{t}^{\mathrm{£}}, \mathrm{s}^{\mathrm{¢}}, \mathrm{q}, \mathrm{x}, \mathrm{y}, \hbar, \mathrm{£} /$ are banned from affixes, cf. (31). Roots on the other hand can include these consonants, as in (32).
(31) Attested and impossible affixes in Muscat Arabic (Glover 1988:153,155,165,199)
a. Pa-

M-
b. ma-

NMLZ
c. -t

1SG.PRF
d. ja-

3PL.IPRF
e. -kum

2PL.M
f. $*^{*}$-s ${ }^{\text {§ }} a x$
g. ${ }^{* *}{ }^{\text {S }}{ }^{\text {aq }}-$
(32) Back and emphatic consonants in roots in Muscat Arabic
a. was $^{\mathrm{q}} \mathrm{ax}$
'to be dirty'
b. $s^{\mathrm{S}} \mathrm{aqat}^{\mathrm{S}}$
'to fall'
c. $\quad$ yammas ${ }^{\text {¢ }}$
'to dip'

There are also languages that have similar but slightly different constraints on roots and affixes. Gooniyandi (Bunaban, Australia) shows such a pattern. In Gooniyandi, roots cannot start with the alveolar tap $/ \mathrm{r} /$ or the palatal lateral $/ \kappa /$. Affixes show a similar yet different restriction: Similar in that they cannot start with a palatal lateral $/ K /$ and different in that they do start with a tap / $/ /$. Examples of possible and impossible morphemes are given in (33). The examples in (33-a,b,c) show attested suffixes beginning with $/ \mathrm{f} /$. The forms in (33-d,e) are impossible affixes with initial $/ \kappa /$. The $/ \kappa /$-initial roots in (33-f,g) are similarly unattested.
(33) Possible and impossible morphemes in Gooniyandi (McGregor 1990:171,226,246)
a. $-[r]$

PL.PRONOUN
b. $\quad-[\mathrm{ri}] \sim[\mathrm{roo}]$
-PAUCAL
c. -[ra]
-MANNER
d. ${ }^{* *}$ - $[\mathrm{Ki}]$
e. **-[Ka]
f. ${ }^{* *}$ raгa
g. ${ }^{* *} \kappa a \kappa$ а

Note that these constraints do not simply refer to syllable structure or word structure. This can be easily seen for the ban on $/ K /$ at the start of an affix. $/ K /$ can occur as the onset of a syllable in morpheme-medial position and thus the ban is not a general constraint on syllables. Examples with root-internal syllable-initial / $\kappa /$ are given in (34). Additionally, it cannot be a restriction on word-initial position, since the suffixes in (33) never occur in word-initial position. Still, they never occur with an initial $/ K /$.
(34) Onset / $K /$ in Gooniyandi (McGregor 1990:142,592,595)
a. baKadi
'flat'
b. ŋaאadi
'testicles'
c. waאara
'sand'

Similarly, the ban on root-initial /r/ cannot be simply a ban on word-initial /r/. Gooniyandi is in fact a weakly prefixing language (cf. Dryer 2013) and roots therefore frequently occur in non-word-initial contexts. ${ }^{7}$ Nevertheless, no root shows an alternation where in a non-initial contexts its initial consonants changes to / $\mathrm{f} /$.

An additional minor finding is that category-specific CoMDs exist but seem to be less frequent than category-indifferent constraints. It seems that category-specific CoMDs are mostly confined to roots. An example of a verb-specific constraint on consonant clusters is given at the end of the next subsection. Further discussion of category-specific effects is provided in chapter 1 and 5 .

Together with the constraints on roots in Ixpantepec Nieves Mixtec and Desano, discussed above, these examples show that constraints can apply to roots or to affixes. This means that the more restrictive hypotheses 2 a and 2 b cannot be upheld. Instead, any theory of phonological constraints on monomorphemic domains has to allow the possibility of constraint referring solely to roots or to affixes.

### 2.3.3 Naturalness of CoMDs

As noted by Albright (2004), constraints on roots (and by extension on other monomorphemic domains) can either be phonetically grounded in articulation or morphologically grounded in parsing. This means that we expect 'unnatural' constraints to occur that cannot be explained by the reduction of articulatory effort.

One especially striking example is a constraint on root-intial consonant in Garifuna (Arawakan; Belize, Honduras, Guatemala, Nicaragua). In Garifuna, the articulatorily unmarked plosive /t/ cannot occur as a root-intial consonant (cf. (35)), even though it is attested in initial position in prefixes and suffixes, cf. (36). This cannot be explained as a contextualized version of a general articulatory constraint on $/ \mathrm{t} /$. Instead, this should be seen as augmentation, i.e. an increase of perceptual markedness, which serves to ease parsing (cf. Smith 2002).

[^15](35) Attested and impossible roots in Garifuna (Haurholm-Larsen 2016:31)
a. ata
'drink'
b. itara
'thus'
c. ${ }^{* *}$ tara
d. ${ }^{* *}$ tija
(36) Attested affixes in Garifuna (Haurholm-Larsen 2016:24)
a. t-
3.F
b. -ti

PST
c. -tija

EMPHATIC__MODALITY
d. $=\mathrm{ti}$

DISCOURSE__ORGANIZATION__CLITIC

Another articulatorily unexpected constraint exists in Ejagham, Fe ' Fe '-Bamileke (AtlanticCongo; Cameroon) and Acoma (Keresan, USA). In these languages, roots can end in phonologically voiced plosives but not in voiceless plosives. Data from Fe'Fe'-Bamileke are presented in (37). Note that the restriction on root-final segments is obscured in word-final position, where final devoicing applies, cf. (37-a). In non-final contexts, the underlying voiced plosive shows up as shown in (37-b). A root with a non-alternating /p/ in both contexts is not attested. (37-a) also shows that there is no general constraint on words that bans final voiceless plosives. (37-d) shows that intervocalic voicing is not a general rule of the language because voiceless plosives can occur intervocalically.

Attested and impossible roots in Fe'fe'-Bamileke (Hyman 1972:50,96)
a. [vap]
/vab/
'to whip'
b. [vabi]
/vab/-/i/
whip-him
'to whip him'
c. ${ }^{* *}[\mathrm{vap}] \sim[$ vapi $]$
d. [toto]
/to/~/to/
ATTR~punch
'(something) to punch'

Again, I will assume that these constraints serve to mark root boundaries in these languages. All three languages prominently feature final devoicing to mark word boundaries. In Ejagham and Fe ' Fe '-Bamileke, plosives become voiceless word-finally. In Acoma, vowels devoice in word-final position. Banning voiceless plosives from root-final position could thus help to distinguish a root-boundary from a word-boundary. This is another instance of a constraint that eases parsing.

Morphologically grounded constraints can also refer to vowel quality. In Mupun (AfroAsiatic; Nigeria), Lango (Nilotic; Uganda) and Ciyao (Atlantic-Congo; Malawi, Mozambique, Tanzania, Zambia), the vowel quality of root/morpheme-initial vowels is restricted. Interestingly, these constraint seem to favor vowels with more articulatory effort. Mupun only allows the low vowel /a/ in root-initial position. Lango prohibits high vowels from starting a root. In Ciyao, roots can only begin with a long vowel.

Data from Lango roots are given in (38). Note that mid and low vowels can occur rootinitially, but high vowels cannot. ${ }^{8}$ This contrasts with the complex words in (39), which start with high vowel prefixes: lax low-toned /ì/- and tense low-toned /ì/.

[^16](38) Attested and impossible roots in Lango (Noonan 1992:30)
a. èllò
'to open, uncover'
b. àmmò
'to handle gently'
c. òllò
'to bore someone'
d. ${ }^{* *} \mathrm{it}$ ḉk
e. **ìnénò
(39) Initial high vowels in complex words in Lango (Noonan 1992:30)
a. ì-tć́k

NC-potato
'potato'
b. ì-nénò

2PL-see
'you saw it'

All of these constraints do not only increase articulatory effort but also perceptibility. This means that they can be seen as an augmentation of the left edge of a root. This facilitates parsing considerably.

Note that this does not mean that all constraints on monomorphemic domains increase the articulatory effort. On the contrary, most CoMDs in the database reduce markedness and could thus also apply in a more general domain. This means that different languages can have opposite constraints in similar domains. A particularity notable example concerns the restriction on consonant clusters in Sye (Austronesian, Vanuatu) and Lango . Whereas consonant clusters are banned from the initial position of verb roots in Sye, they are restricted to that very same root-initial position in Lango. Thus, in Sye, the roots in $(40-\mathrm{a}, \mathrm{b})$ are possible noun roots but not possible verb roots. Consonant clusters in medial position are allowed in verbs (40-c,d).
(40) Consonant clusters in roots in Sye (Crowley 1998:20,22)
a. /nvat/
'stone'
b. /pvuvyum/
'morning'
c. /jahnror/-
'to pull out'
d. /empyu/-
'to dance'

In Lango on the other hand, consonant clusters are restricted to root-intial position, as shown in (41). This is not a constraint on words, since forms with partial reduplication can violate this constraint, as shown in (42). It is evident that the two constraints in Sye and Lango impose opposite requirements.
(41) Attested and impossible roots in Lango (Noonan 1992:7)
a. kwèè
'to cool'
b. bjè
'termite hill'
c. **tùtwàl
d. **nwénwé
(42) Consonant clusters in derived non-initial syllables in Lango (Noonan 1992:114)
a. tù~twàl
much~very
'very much
b. nwé~nnwé
sort.of~smelly
'sort of smelly'
c. dwó~dwôy
sort.of~big
'sort of big'

All in all, CoMDs can thus either be motivated by morphological parsing or by the reduction of articulatory effort. This is in accordance with the first part of hypothesis 3. CoMDs are thus potentially different from more general markedness constraints.

### 2.3.4 Discussion of Exploratory Results

Three conclusions can be drawn from this exploratory study. First, CoMDs can make reference to derived properties like prosodic structure and contextual allophones. Second, CoMDs can take roots as well as affixes as their domain. Third, CoMDs can be motivated by articulatory phonetics as well as morphological parsing. Especially the first two conclusions have consequences for a theory of phonological constraints on monomorphemic domains.

The first two conclusions clearly falsify the hypotheses $1,2 \mathrm{a}$ and 2 b discussed in section 2.1. This also has consequences for a possible theory of CoMDs. The reference to prosodic structure and contextual allophones requires a theory that makes reference to surface structure in the output, such as parallel Optimality Theory. A theory based on CoURs, which only restricts underlying representations, is not sufficient to account for these patterns.

Similarly, CoMDs clearly refer to roots and affixes as separate domains. This means that the Root-Affix Faithfulness Metaconstraint by McCarthy \& Prince (1995) is not sufficient to explain all attested CoMDs and neither are strictly cyclic theories. The problem is that the Root-Affix Faithfulness Metaconstraint only allows affix-specific CoMDs and strictly cyclic theories do not allow phonological generalizations to hold for bound forms. It has, however, been noted before that constraints on affixes are more likely to be due to chance because affixes form a closed class (cf. e.g. McCarthy \& Prince 1995:fn. 93) and often occur in prosodically non-prominent positions. In order to exclude chance as an explanation, an inventory of possible phonological sequences and the total number of affixes have to be used to compare the expected and attested incidence of certain segments. Such a statistical analysis for individual languages depends on the data available and is not feasible within this dissertation. One possible case study for future research might be Tirmaga (Surmic; Ethiopia, South Sudan), where the constraint on affixes is stricter than in Muscat Arabic. All consonant-initial suffixes have to begin with a coronal consonant (Bryant 1999). In section 2.5, I will show that the theory proposed in this dissertation, Hierarchical Morphoprosodic Structure, can account for all of these domains.

### 2.4 Inferential Results

In this section, I will provide more detailed inferential results of the typological study. I will show that the variables are not independent in the g -sample extracted from DoCoMD. The main hypothesis that I will test assumes that certain values of constraint properties cluster together. As stated in Hypothesis 3, morphological parsing is facilitated by positional and sequential constraints on root consonants. These constraints should therefore be more common. Similarly, I will show that inventory constraints are more often restricted to
vowels in non-roots.
(43) Clusters of constraint property values

$$
\begin{array}{ll}
\text { inventory constraints } & \leftrightarrow \text { positional \& sequential constraints } \\
\text { non-root constraints } & \leftrightarrow \text { root constraints } \\
\text { vowel constraints } & \leftrightarrow
\end{array} \text { consonant constraints } ~ \$ ~ l
$$

I will, however, also show that these tendencies might be influenced by areal effects. As shown in subsection 2.4.2, vowel and inventory constraints are more common in Africa, whereas sequential constraints and constraints on consonants are more frequent in Papuanesia. This suggests that alternative explanations are also possible. I discuss these in subsection 2.4.3.

As evident from the description above, I will use cover values in some cases. The category root constraints vs. non-root constraints deserves further explanation. Root constraints include both the value root and root proper for the domain property. Non-root constraint on the other hand conflates the values morpheme and affix.

The purpose of this reduction of categories is to ease statistical analysis. As suggested by Janssen et al. (2006), $\chi^{2}$-tests will be used, since all data are of the nominal data type. Janssen et al. (2006) state that other well-known statistical tests like ANOVA and t-tests should not be applied to genealogically controlled samples, since they presuppose a randomly selected sample. The use of $2 \times 2$ contingency tables allows the expected values to be above 5 and therefore randomization tests are not needed (cf. Janssen et al. 2006).

### 2.4.1 Interaction of Properties

Starting with the interaction of constraint type and constraint domain, the table in (44) shows the frequency of positional constraints and root domains. ${ }^{9}$ Positional root constraints $(\mathrm{n}=49)$ are far more common than positional constraints on non-roots ( $\mathrm{n}=22$ ). Note that for non-positional constraints, the distribution is more balanced ( $\mathrm{n}=64 \mathrm{vs} . \mathrm{n}=65$ ).
(44) Positional constraints are more common in roots $\left(\chi^{2}=7.0139 ; \mathrm{p}=.008088\right)$

|  | root | non-root | total |
| :--- | ---: | ---: | ---: |
| Positional | 49 | 22 | 71 |
| Non-Positional | 64 | 65 | 129 |
| total | 113 | 87 | 200 |

Nevertheless, a converse generalization is possible. Inventory constraints are more common

[^17]in non-roots. As shown in the table in (45), there are only $n=7$ inventory constraints in the whole g-sample. This starkly contrasts with the high number of non-inventory constraints in roots $(\mathrm{n}=106)$. Note again that the sample is more balanced for non-roots ( $\mathrm{n}=40 \mathrm{vs}$. $\mathrm{n}=47$ ).
(45) Inventory constraints are more common in non-roots $\left(\chi^{2}=43.2732 ; \mathrm{p}<.00001\right)$

|  | root | non-root | total |
| :--- | ---: | ---: | ---: |
| Inventory | 7 | 40 | 47 |
| Non-Inventory | 106 | 47 | 153 |
| total | 113 | 87 | 200 |

Let us now turn to the interaction between substance and domain. The table in (46) shows the interaction for consonant vs. non-consonant constraints. The number of non-consonant constraints in roots is significantly smaller than expected ( $\mathrm{n}=46$ ). Similar, consonant constraints in roots are underattested ( $\mathrm{n}=30$ ).

Consonant constraints are more common in roots $\left(\chi^{2}=12.1129 ; \mathrm{p}=.000501\right)$

|  | Root | Non-Root | total |
| :--- | ---: | ---: | ---: |
| Consonant | 67 | 30 | 97 |
| Non-Consonant | 46 | 57 | 103 |
| total | 113 | 87 | 200 |

Again, the converse observation is also significant. Vowel constraints are more common in non-roots, as illustrated in the table in (47). ${ }^{10}$ Root constraints only very rarely affect vowels ( $\mathrm{n}=16$ ). Inside non-roots on the other hand, vowel constraints are significantly more common than expected ( $\mathrm{n}=47$ ).

Vowel constraints are more common in non-roots ( $\chi^{2}=36.2011$; $\mathrm{p}<0.00001$ )

|  | Root | Non-Root | total |
| :--- | ---: | ---: | ---: |
| Vowel | 16 | 47 | 63 |
| Non-Vowel | 97 | 40 | 137 |
| total | 113 | 87 | 200 |

There is also a significant interaction between substance and constraint type. The table in (48) shows that non-inventory constraints are significantly less common for vowels ( $\mathrm{n}=28$ ). Similarly, inventory constraints are less frequent than expected for non-vowel substance

[^18]( $\mathrm{n}=12$ ).
(48) Inventory constraints are more common on vowels $\left(\chi^{2}=52.5687\right.$; $\left.\mathrm{p}<0.00001\right)$

|  | Vowel | Non-Vowel | total |
| :--- | ---: | ---: | ---: |
| Inventory | 35 | 12 | 47 |
| Non-Inventory | 28 | 125 | 153 |
| total | 63 | 137 | 200 |

Again, there is a converse result for consonants. Sequential constraints are more common in consonants, cf. the table in (49). The distribution of sequential constraints on consonants is unexpected. Sequential constraints on non-consonants ( $\mathrm{n}=22$ ) are significantly less common than constraints on consonants ( $\mathrm{n}=47$ ).

| Sequential constraints are more common on consonants |  |  |  |
| :--- | ---: | ---: | ---: |
|  | Consonant | Non-Consonant | total |$\left.\chi^{2}=16.2285 ; \mathrm{p}=.000056.\right)$

All in all, we can summarize the inferential results as follows. The properties of a non-root domain, vowel substance and inventory type constraint cooccur significantly more often than expected. Similar statements are true for the following pairs of values: root domain and positional type, root domain and consonant substance, sequential type and consonant substance.

### 2.4.2 Areal Distribution

The areal effects seem to mirror the interaction of the different categories. This becomes evident when looking at the three best-represented macro-areas, Africa, Eurasia and Papuanesia. Whereas Eurasia shows no skewing with respect to constraint properties, this is not true for Africa and Papuanesia. In Africa, vowel inventory constraints are more common. In Papuanesia, sequential constraints on root consonants are more frequent. All of these results should be taken with a grain of salt since I have no evidence for the absence of certain constraint types or properties in certain areas or families. Additionally, a sampling effect cannot be completely ruled out.

Starting with the constraint type, the table for sequential constraints in Papuanesia is presented in (50). In Papuanesia, sequential constraints are significantly more common than expected. Non-sequential constraint on the other hand are significantly less common than elsewhere.
(50) Sequential constraints are more common in Papuanesia $\left(\chi^{2}=30.9869 ; \mathrm{p}<0.00001\right)$

|  | Papuanesia | Non-Papuanesia | total |
| :--- | ---: | ---: | ---: |
| Sequence | 33 | 36 | 69 |
| Non-Sequence | 16 | 115 | 131 |
| total | 49 | 151 | 200 |

Papuanesia also shows a skewed distribution of constraints on consonants. These constraints are overrepresented in Papuanesia ( $\mathrm{n}=39$ ), whereas constraints on non-consonants are underrepresented $(\mathrm{n}=10)$. The results are again significant.

Constraints on consonants are more common in Papuanesia $\left(\chi^{2}=25.1185 ; \mathrm{p}<0.00001\right)$

|  | Papuanesia | Non-Papuanesia | total |
| :--- | ---: | ---: | ---: |
| Consonant | 39 | 58 | 97 |
| Non-Consonant | 10 | 93 | 103 |
| total | 49 | 151 | 200 |

Finally, constraint domains are also not equally distributed in Papuanesia. Root constraints are significantly more common in Papuanesia ( $\mathrm{n}=37$ ). Non-root constraints on the other hand are less frequent than expected ( $\mathrm{n}=12$.)
(52) Root constraints are more common in Papuanesia $\left(\chi^{2}=9.543 ; \mathrm{p}=.002007\right)$

|  | root | non-root | total |
| :--- | ---: | ---: | ---: |
| Papuanesia | 37 | 12 | 49 |
| Non-Papuanesia | 76 | 75 | 151 |
| total | 113 | 87 | 200 |

Another macro-area with a unbalanced distribution of some constraint properties is Africa. The table in (53) shows the number of constraints on vowels and non-vowels in Africa and elsewhere. In Africa, vowel domain constraints $(\mathrm{n}=40)$ are more common than in the rest of the world $(\mathrm{n}=23)$. Note that this is not only significant, but it also shows that there are more vowel constraints in Africa than all other macro-areas combined in the g-sample.
(53) Vowel constraints are more common in Africa ( $\chi^{2}=23.1903 ; \mathrm{p}<0.00001$ )

|  | Africa | Non-Africa | total |
| :--- | ---: | ---: | ---: |
| Vowel | 40 | 23 | 63 |
| Non-Vowel | 38 | 99 | 137 |
| total | 78 | 122 | 200 |

Africa also shows a preference for a certain constraint type. As shown in the table in (54), inventory constraints are more common than expected in Africa ( $\mathrm{n}=32$ ). Again, this number is higher than in all other macro-areas combined ( $\mathrm{n}=15$ ), where inventory constraints are significantly less common.

| Inventory constraints are more common in Africa $\left(\chi^{2}=21.8466 ; \mathrm{p}<0.00001\right)$ |  |  |  |
| :--- | ---: | ---: | ---: |
|  | Africa | Non-Africa | total |
| Inventory | 32 | 15 | 47 |
| Non-Inventory | 46 | 107 | 153 |
| total | 78 | 122 | 200 |

We can summarize the areal distrubution in the following way. In Papuanesia, sequential constraints on root consonants are more common. In Africa on the other hand, vowel inventory constraints are more frequently attested.

### 2.4.3 Discussion of Inferential Results

The inferential results on the interaction of constraint properties seem to be very clear. Vowel inventory constraints tend to apply in non-roots, sequential and positional constraints tend to apply to root consonants. However, they should be interpreted with caution and only serve to formulate more precise hypotheses on the correlation of CoMDs. The very skewed areal distribution suggests that alternative explanations of the results are possible. They could also be explained by areal or other typological effects. I will discuss the three possible explanations one-by-one.

First, it could be true that these correlations are real in the overall number of languages and the effects are just stronger in certain macro-areas. This would mean that hypothesis 3 is partially true. These effects can then potentally be related to morphological parsing.

Second, the areal effects could be primary, such that a larger, areally controlled sample would find no effects for the whole sample but would replicate the strong areal effects. This would mean that there is, for instance, a correlation between constraints on vowels and inventory constraints in Africa but not elsewhere. Similarly, the strong correlation of sequential constraints on root consonants could be restricted to Papuanesia and be absent
in other areas. The size of the present g -sample does not allow us to test this hypothesis.
Third, it could be true that the present $g$-sample shows a typological bias. The g-sample extracted from DoCoMD includes a high amount of languages with infixation and vowel harmony. It is conceivable that constraints on consonantal sequences are more common in languages with infixes or that vowel inventory constraints are more common in languages with vowel harmony. If this is true, it should be possible to test this hypothesis with a larger sample that specifically tests or controls this factor. Such a study is beyond the scope of this thesis.

We have thus seen that the correlation between different constraint properties seeming to favor vowel inventory constraints on non-roots and sequential/positional constraint on root consonants. These correlation could be explained by morphological parsing requirements. Areal skewing, however, suggests that alternative explanation based on areal effects and typological bias should also be considered. I leave this question open for future research.

### 2.5 General Discussion

In this section, I will go on to show how the exploratory results can be accounted for in an analysis based on hierarchical prosodic structure. This will be followed by a discussion of alternative approaches to the phonology-morphology interface and the problems that they face. A more detailed summary of all arguments provided in this dissertation is postponed to chapter 6 .

The first conclusion drawn from the exploratory study was that CoMDs need to refer to derived properties such as syllables, moras and contextual allophones. This is entirely unproblematic in a theory based on Morphoprosodic Hierarchical Structure proposed here. Since CoMDs are modeled as markedness constraints in an Optimality Theory framework, they can only refer to surface structure, including any derived structure such as syllable structure, moras and contextual allophones. The constraints in (55), for example, are sufficient to derive the allophonic pattern in Desano. Recall that in Desano [d] is banned from root-medial position, where its positional allophone [r] occurs. These constraint cannot be easily formulated as a constraint on underlying presentations, since $[r]$ is not present underlyingly.
(55) Constraints needed for Desano CoMDs on contextual allophones
a. $\quad[\sqrt{\omega} \mathrm{d}$

Count one violation for every [d] that does not occur in root-initial position.
b. ${ }^{*}{ }_{r}$

Count one violation for every $[r]$.

The second conclusion from the exploratory study concerned the availability of different types of domains for CoMDs. Affixes and roots are both possible domains for phonological generalizations. Again, this can be derived via Hierarchical Morphoprosodic Structure. Consider the adjunctive, binary branching prosodic structure in (56). A prosodic word dominates only the root. Another complex prosodic word recursively dominates this simple prosidic word and the closer affix. Yet another prosodic word dominates this complex prosodic word and the outer affix. This structure will allow us to formulate three different domains to which constraints can be relativized. Note that this structure involves 'recursive' prosodic words. The recursivity of prosodic constituents is a hotly debated topic, cf. Itô \& Mester $(2012,2021)$ for arguments in favor of recursive prosody and Downing \& Kadenge (2020) and Miller \& Sande (2021) for recent arguments against word-level recursion. I do not provide direct empirical evidence for recursive structure here. Instead, I show that this assumption allows us to derive monomorphemic affix domains from prosodic structure.

Adjunctive, binary branching structure


In order to define the correct domains, we need to distinguish between the different prosodic words. As proposed in chapter 1, I will distinguish between non-minimal prosodic words ( $\omega$, NMIN ) and prosodic root words. A non-minimal prosodic word is any prosodic word that dominates at least one other prosodic word. The minimal prosodic word that dominates a root on the other hand, will be termed the prosodic root word $(\sqrt{\omega})$. Note that these notions are not primitives of the theory but derived from the simpler notions of domination and root status. The different prosodic words are marked in the tree in (57).
(57) Different kinds of prosodic words


From these different kinds of prosodic words, we can derive different domains. Both affixes have in common that they are directly dominated by their own non-minimal prosodic word. ${ }^{11}$ Roots on the other hand constitute the only phonological material that is dominated by the prosodic root word. This means that Hierarchical Morphoprosodic Structure allows us to differentiate between roots and affixes. It also allows us to refer to monomorphemic domains in general. This can be accomplished by referring to any material that is directly dominated $\left(=_{\mathrm{DD}}\right)$ by any prosodic word. Some examples for constraints with different domains are given in (58). The constraint in (58-a) restricts [d] to the initial position in a prosodic root word, as needed in Desano roots. In (58-b), a constraint for the analysis of Muscat Arabic is given. Back and emphatic consonants are prohibited from occurring if directly dominated by a non-minimal prosodic word. This effectively penalizes such consonants in affixes. The constraint in (58-c) is violated by any sequence of three syllables that is directly dominated by the same prosodic word, i.e. belonging to the same morpheme. See also chapter 1 and chapter 5 for category-specific CoMDs and their analysis by Prosodic Category Inheritance using Hierarchical Morphoprosodic Structure.

[^19](58) Some examples of CoMDs as constraints relativized to morphoprosodic domains
a. $[\sqrt{\omega} \mathrm{d}$

Count one violation for every [d] that does not occur in the initial position of a prosodic root word (cf. Desano roots).
b. ${ }^{*} \mathrm{Q}_{\omega, \mathrm{NMIN}, \mathrm{DD}}$

Count one violation for every back or emphatic consonant $(=\mathrm{Q})$ that is directly dominated by a non-minimal prosodic word (cf. Muscat Arabic affixes).
c. ${ }^{*} \sigma \sigma \sigma_{\omega, \mathrm{DD}}$

Count one violation for every prosodic word that directly dominates three syllables (cf. Ejagham morphemes).

As for the third conclusion on the naturalness of CoMDs, I would like to note that not all possible constraints generated by a constraint scheme must be part of the CON, the universal constraint set in Optimality Theory. This means that not every constraint that exists must also exist in versions that are relativized to every possible domain. Constraints do not even have to exist in a general version. This can be achieved by understanding grounding in phonetics and parsing as filters that act upon the set of all possible domains. Such approaches have been proposed by Smith (2002) and Bermúdez-Otero \& Börjars (2006) among others.

Let us now turn to alternative proposals. One way to account for CoMDs is to restrict the underlying forms by specific constraints, ie. CoURs. This is highly problematic given the general empirical picture. CoMDs can refer to derived prosodic structure. Recall, for example, that morphemes in Ejagham cannot include more than two syllables. Since syllables are predictable, they are usually assumed to be derived and not present in the underlying representation. However, if a CoUR needs to be formulated on syllables for Ejagham morphemes, syllables need to be present in the underlying form. This defies the purpose of underlying representations that are supposed to exclude any predictable information. CoMDs on derived phonological problems thus pose a problem for any theory that derives all CoMDs as constraints on underlying representations.

Strictly cyclic theories, such as Cophonology Theory (Orgun 1996; Inkelas 1998), face a different challenge. In cophonology theory, potentially different phonological grammars $\left(\phi_{i}\right)$ apply at every affixation step or to every morphological construction respectively, as shown in the structure in (59). This means that even though there can be an affix-specific phonological grammar, affixes cannot be a domain for phonological computation. As soon as they are subject to a phonological grammar, they enter a larger domain. Therefore, they can never be their own domain, as far as phonology is concerned. This is problematic, as we have seen that there are constraints that require affixes to form a domain for constraints
on monomorphemic domains. The same challenge is faced by OT-OO and Stratal OT without a root stratum.

Cyclic application of cophonologies in Cophonology Theory


The opposite problem occurs if we assume the Affix-Root Faithfulness Metaconstraint. This metaconstraint states that, in an Optimality Theory constraint ranking, affix-specific faithfulness should always be outranked by root-specific faithfulness constraints. This is problematic for CoMDs that only restrict roots. Consider the case of Garifuna / $\mathrm{t} /$, which only occurs at the left edge of affixes but not at the left edge of roots. Assuming that there can be a more general markedness constraint against morpheme-initial [ t ], i.e. * $[\mathrm{t}$, we can only get three possible rankings. If the markedness constraint outranks all faithfulness constraints, we get a more general pattern, without any affix or root-specific effects. Similarly, the markedness constraint becomes inactive if it is ranked below all faithfulness constraints. The only relevant possible ranking is Faith-Root $\gg$ *[ $\mathrm{t} \gg$ FaithAffix. However, as shown in the tableau in (60), this ranking predicts the absence of affix-initial $/ \mathrm{t} /$, contrary to fact. ${ }^{12}$ The correct candidate ( $60-\mathrm{c}$ ) with affix inital $/ \mathrm{t} /$ but not root-initial / t / is excluded since it violates the high ranked root-faithfulness constraint. The root-intial /t/ can never be deleted without the affix-intial /t/ also deleting.
(60) Affix-Root Faithfulness Metaconstraint fails to predict Garifuna / t / root constraint

| I: | $\mathrm{t}_{\mathrm{R}} \mathrm{a}_{\mathrm{R}} \mathrm{t}_{\mathrm{R}} \mathrm{a}_{\mathrm{R}}$-tija | FAITH-ROOT | $*[\mathrm{t}$ | FAITH-AFFIX |
| ---: | :---: | :---: | :---: | :---: |
| a. | $\mathrm{t}_{\mathrm{R}} \mathrm{a}_{\mathrm{R}} \mathrm{t}_{\mathrm{R}} \mathrm{a}_{\mathrm{R}}$-tija |  | $* *!$ |  |
| -b. | $\mathrm{t}_{\mathrm{R}} \mathrm{a}_{\mathrm{R}} \mathrm{t}_{\mathrm{R}} a_{\mathrm{R}}$ - ija |  | $*$ | $*$ |
| erc. | $\mathrm{a}_{\mathrm{R}} \mathrm{t}_{\mathrm{R}} \mathrm{a}_{\mathrm{R}}$ - tija | $*!$ | $*$ |  |

The Affix-Root Faithfulness Metaconstraint faces a further minor challenge. Without proper representations, it cannot refer to monomorphemic domains at all. If any phonological material is solely marked for its affix or root status, we would expect adjacent affixes to form one domain. This would make it impossible to formulate constraints on individual affix domains. Each affix has to be at least marked for being different from adjacent affixes

[^20]either through a boundary marker or through coding of morphological identity, cf. chapter 1 and chapter 3 for further discussion.

### 2.6 Conclusion

In this chapter, I have provided an overview of the crosslinguistic picture with regard to phonological constraints on monomorphemic constraints. The exploratory study has shown that there is a wider variety of constraints than hypothesized before. They can refer to underlying phonological properties as well as to derived properties. Their domain can be either affixes, roots or both. Additionally, they can be grounded either in articulatory phonology or in morphological parsing.

The first two findings pose difficult challenges to several theories of the phonologymorphology interface. Theories that rely on constraints on underlying representations to derive CoMDs face considerable problems when dealing with CoMDs on derived phonological properties. Strictly cyclic theories and the Root-Affix Faithfulness Metaconstraint face problems with CoMDs on affixes and roots respectively. Both theories predict a more restricted set of possible domains.

The same findings are entirely unproblematic for a Hierarchical Morphoprosodic Structure approach. Markedness constraints, relativized to prosodic domains, apply to the output and therefore have access to derived phonological properties. Similarly, direct domination by non-minimal prosodic words and prosodic root words allows us to define monomorphemic affix and root domains, respectively.

Additionally, I have shown that in a genealogically controlled sample, certain properties of CoMDs seem to cluster. Inside non-roots, we find more vowel inventory constraints than expected. Positional constraints and constraints on consonants on the other hand are more frequent in roots. This could be seen as an effect of the requirements of morphological parsing. However, I have also provided two alternative explanations, either as areal effects or as typological bias introduced by languages with infixes and vowel harmony.

## Chapter 3

## The Domain Problem in Infixation

As argued in chapter 1, a crucial aspect of Morpheme Structure Domains is their ability to apply exclusively to monomorphemic domains. Judging from the results of the crosslinguistic survey in chapter 2 , it is clear that restrictions on monomorphemic domains are empirically real. Kisseberth \& Kenstowicz (1977) refer to this problem as the Domain Problem, cf. (1).
(1) The Domain Problem (Kisseberth \& Kenstowicz 1977:149)

Some constraints seem to refer to the morpheme as it appears in the lexicon, other constraints refer to the structure that exists after words have been formed. These constraints refer either to the word or to the syllables that comprise the word.

The Domain Problem poses a challenge for Optimality Theory with Richness of the Base and some degree of Output-Drivenness and modularity. Richness of the Base excludes the possibility of referring to monomorphemic domains before a derivation using morphological concatenation and phonological processes applies (cf. chapter 1). Output-Drivenness (cf. Tesar 2013) bans constraints that refer to information on morphological domains that are present only in the input. Finally, modularity reduces the amount of accessible morphological information in the output of the phonological computation. Nevertheless, several approaches have been developed to allow minimal reference to morphological information in Optimality Theory by introducing representational or procedural devices.

In this chapter, I will compare these devices with regards to a special case of monomorphemic domains: non-contiguous output domains created by infixation. Even though all of the proposed theoretical devices can derive simple monomorphemic domains, I will argue that the empirical picture on non-contiguous domains allows us to distinguish their predictions. I will show that of four expected possible interactions between constraints
on monomorphemic domains and infixation, only two are attested. An infix can either become part of the root domain or not, but a root never loses its status as a domain through infixation. In other words, a discontiguous root is always subject to root domain constraints. This empirical picture, mainly based on Austronesian and Semitic languages, is best analyzed using Hierarchical Morphoprosodic Structure coupled with a fixed ranking of markedness constraints relativized to a narrow domain above markedness constraints relativized to a wider domain. I will show that procedural approaches either overgenerate because they allow for infixation to bleed the application of a root domain constraint to the discontiguous root or undergenerate because they force the infix to become part of the root domain. A representational approach on the other hand can be easily restricted to only account for the attested patterns. The main reason for the success of such an approach is that infixes are representationally special.

Non-contiguous output domains are a suitable testing ground for the Domain Problem because they show a mismatch between input and output domains. In the respective lexical underlying representations, two domains will be contiguous and separate. In the output, however, the domains will be intertwined. Since one domain is embedded in the other domain, the outer domain becomes non-contiguous. The inner domain intervenes between the string preceding it and the string following it. These non-contiguous domains can come about by morphological operations, i.e. infixing. This means that potentially there are different possibilities for output domains. A strictly output string-based parsing of the domains could include three domains, a pre-infix root domain, given as [ro] in (2-b), followed by the infix domain, and a second root domain - a post-infix root domain [ot]. Note that such a parse would yield three contiguous domains without any embedding or discontiguity. Alternatively, if an analysis either allows reference to the input domains (cf. (2-a)) or allows for nested/non-contiguous domains, the parse shown in (2-c) is possible. The root is still a simple domain just as it was in the input. It is however split up by the infix domain. Finally, the input domains could fuse into a single domain in the output, as shown in (2-d). In this chapter, I will show that a nested domain (2-c) is needed, because split domains and single domains predict unattested empirical patterns.
(2) Possible Output Domains for Infixation
a. Input: [root] [infix]
b. Split Domain Output: [ro][infix][ot]
c. Nested Domain Output: [ro[infix]ot]
d. Single Domain Output: [ro infix ot]

Infixes are characterized by what Yu (2007:9) calls Derived Discontinuity. An infix will split up the contiguous base morpheme. This potentially creates a non-contiguous
monomorphemic domain，e．g．a non－contiguous root．This becomes relevant if a constraint on monomorphemic domain applies to roots．The following sections will deal with such root domain constraints，that restrict certain consonant cooccurrences in roots．The main questions will be the following：Do the infix and the root form a joint domain for the application of the constraint？Relatedly，does the root still form a domain for the constraint when infixed？The first question will be answered on a language－specific basis， whereas the second question can be universally answered negatively．The generalization that emerges states that a complex root＋infix domain is only available if the root still forms a domain in the same infixed form．In the following two sections 3.1 and 3.2 ，I will present data from two different languages that exemplify both attested patterns，before introducing the Hierarchical Morphoprosodic Structure analysis in section 3．4．I will show that Hierarchical Morphoprosodic Structure can be easily and naturally restricted to only predict attested patterns．The following two sections will argue against procedural approaches and against simpler representations．After discussing more crosslinguistic data in section 3．7，I will conclude in section 3．8．

## 3．1 Muna Labial Infix Dissimilation

Many Austronesian languages show root consonant cooccurrence restrictions similar to the well－known patterns of Semitic languages．Additionally，many of them feature an infix $\langle u m\rangle$ or a cognate．In Muna（Austronesian，Indonesia），$\langle u m\rangle$ is used in the irrealis form of a verb．It has alternant forms，which according to Pater（2001）are used to avoid violating an OCP－like constraint on labial non－identical consonants in the root domain．The Muna〈um〉 infix is regularly attached into the stem after the first consonant（cf．van den Berg 1989）．
（3）Muna regular infixation（van den Berg 1989：32）
a．dadi， $\mathrm{d}\langle\mathrm{um}\rangle$ adi
live $\langle\mathrm{IRR}\rangle$ live
b．dudu，d $\langle$ um $\rangle$ udu
push 〈IRR〉push
c．gaa，g（um）aa
marry $\langle\mathrm{IRR}\rangle$ marry

Muna has a set of restrictions on possible consonants in roots．One of them states that the bilabial nasal $/ \mathrm{m} /$ cannot occur as the second consonant of a root if the first consonant is a labial obstruent．In a corpus of 1,100 disyllabic roots，van den Berg（1989）finds
no occurrences of medial $/ \mathrm{m} /$ following initial labial obstruents. ${ }^{1}$ I will refer to this generalization as the *BM-constraint. In derived forms however, such sequences are allowed, cf. (5).
(4) Muna impossible roots
a. ${ }^{* *}$ pimi
b. ${ }^{* *}$ femo
c. ${ }^{* *}$ bumo
d. ${ }^{* * m}$ boma
e. **6ama
(5) Muna labial obstruent followed by $/ \mathrm{m} /$ in morphologically derived environments
a. do-po-mai-nsule

3PL.REA-RECP-come-return
'they go back and forth between home and work' (van den Berg 1989:330)
b. no-fo-me-owa-ghoo foo

3SG.REA-DETRANS-INFLECTION.CLASS-bring-INDIRECT.OBJECT mango
'he brought us a mango' (van den Berg 1989:72)
c. no-rubu-mo

3SG.REA-small-PRF
'It became small' (van den Berg 1989:143)
d. Ge-mo anoa
with-PRF 3SG
'with him' (van den Berg 1989:307)
e. $\quad \mathrm{ra}^{\mathrm{m}} \mathrm{bi}$-mo
blow-PRF
'o' clock' (van den Berg 1989:334)
f. $\quad a^{m} \mathbf{p a - m o}$
merely-PRF
'only' (van den Berg 1989:307)

The /um/-infix is affected by the *BM-constraint. It cannot be simply infixed into a root that begins with a labial obstruent. There are two possible repair strategies. With roots beginning in $/ \mathrm{f} /$ and $/ \mathrm{p} /$, nasal accretion applies, i.e. the root-initial consonant is substituted by $/ \mathrm{m} /$ instead of attaching the infix regularly. ${ }^{2}$

[^21](6) Muna exceptional nasal accretion (van den Berg 1989:32)
a. $\quad \mathrm{po}^{\mathrm{g}} \mathrm{ko}, \mathrm{m}$-ongko, ${ }^{*} \mathrm{p}\langle\mathrm{um}\rangle \mathrm{o}^{\mathrm{g}} \mathrm{ko}$
kill IRR-kill
b. pili, m-ili, $\quad{ }^{*}$ p $\langle u m\rangle$ ili
choose IRR-choose
c. foni, m-oni, $*_{f}\langle u m\rangle$ oni climb IRR-climb
d. futaa, m-utaa, $*_{f}\langle u m\rangle$ utaa laugh IRR-laugh

With roots beginning in $/ \mathrm{b} /, / \mathrm{b} /, /{ }^{\mathrm{m}} \mathrm{b} /, /{ }^{\mathrm{m}} \mathrm{p} /$, the infix exponent is deleted altogether. ${ }^{3}$ Again, this is a further repair strategy for the ban on labial obstruents preceding $/ \mathrm{m} /$, the *BM-constraint. The distinction between the two repair strategies and its possible causes will not further concern us here.
(7) Muna exceptional infix deletion (van den Berg 1989:32)
a. baru, baru, $\quad{ }^{*}\langle\langle u m\rangle a r u$
happy happy[IRR]
b. Gala, Gala, * $6\langle$ um $)$ ala
big $\operatorname{big}[$ IRR $]$
c. ${ }^{\mathrm{m}}$ bolaku, ${ }^{\mathrm{m}}$ bolaku, ${ }^{* \mathrm{~m}} \mathrm{~b}\langle u m\rangle$ olaku
steal steal[IRR]

If the root would cease to be a domain for the *BM-constraint, we would expect the forms in (8), contrary to fact. A nasal shows up in these forms as the second consonant that did not show up in a non-infixed root. This nasal is present in the underlying representation but dissimilated in cases without infixation. This means that the root domain constraint would only apply to contiguous roots, i.e. roots that did not undergo infixation. In infixed forms, such a repair mechanism would be blocked. A more detailed discussion of such a pattern will be postponed to section 3.3.
(8) Unattested root alternations in Muna
a. $\left[\mathrm{po}^{\mathrm{y}} \mathrm{ko}\right] \sim^{* *}[\mathrm{~m}-\mathrm{omo}]$ from underlying /pomo/
b. [foni] $\sim^{* *}[\mathrm{~m}$-omi] from underlying /fomi/

The main observation here is that the infix and the root form a domain for the prohibition

[^22]against /m/ following labial obstruents. Two different repairs apply if the *BM-constraint would be violated in the root+infix domain. The root, however, still forms a domain, since no violations of the *BM-constraint are found in roots inside infixed forms, cf. the examples in (8).

### 3.2 Hebrew Hitpa'el Metathesis

Semitic reflexive forms often include an exponent /t/ that is variably infixed or prefixed. In Modern Hebrew (Afro-Asiatic, Israel) these are combined with a stable prefix /hi/-. ${ }^{4}$ This pattern bears on the question of non-contiguous domains.

In Hebrew, like in many other Semitic languages, triconsonantal roots cannot include two identical consonants as two non-final adjacent members (cf. McCarthy 1986:209, Bachra 2001:14). I will refer to this generalization as $\operatorname{OCP}(\mathrm{C}) .{ }^{5}$
(9) $\operatorname{OCP}(\mathrm{C})$ in Hebrew roots
a. katav, katv-u
write.PST write.PST-3PL
'he wrote, they wrote'
b. ${ }^{* *}$ kakav, ${ }^{* *}$ kakvu
c. ${ }^{* *}$ tatav, ${ }^{* *}$ tatvu
d. namax, minef
become.short.PST leverage.PST
'he became shorter, he leveraged'
e. ${ }^{* *}$ mamax, ${ }^{* *}$ nanax, ${ }^{* *}$ mimef, ${ }^{* *}$ ninef
f. ${ }^{* *}$ mamxu, ${ }^{* *}$ nanxu, ${ }^{* *}$ mimfu, ${ }^{* *}$ ninfu

This is not a general restriction on words. The first two consonants in a word can be identical if one of them is part of a prefix, cf. (10). The OCP(C) thus does not hold in derived words.

[^23](10) No $\operatorname{OCP}(\mathrm{C})$ in Hebrew derived words
a. m-amxiz

PTCP-dramatize
'he dramatizes (a play)' (cf. **mamxu)
b. t-atxil

2M.FUT-start
'you (masc) will start' (cf. **tatvu)
c. n-anxil

1PL.FUT-endow
'we will endow' (cf. ** nanxu)

The reflexive form of a verb, also known as Hitpa' 'el, usually includes a prefixal exponent /hit/-. The meaning of these forms is often not fully transparent. Examples are given in (11).
(11) Reflexive prefix in Hebrew
a. hit-xamek

REFL-evaded
'he evaded'
b. hit-batel

REFL-cancel
'he was canceled'

In a specific phonological context, however, /t/ is infixed. If the first consonant of the root is a sibilant, $\langle\mathrm{t}\rangle$ occurs directly after it. ${ }^{6}$ In example (12-a), $\langle\mathrm{t}\rangle$ occurs after $/ \mathrm{s} /$ and in (12-b) it occurs after $/ \mathrm{J} /$. In both cases, infixation applies.
(12) Reflexive infix in Hebrew
a. hi-s $\langle t\rangle$ arek,
$\mathrm{REFL}_{i}-\left\langle\mathrm{REFL}_{i}\right\rangle \mathrm{Comb}$
'he combed himself'
b. hi- $\int\langle\mathrm{t}\rangle$ alev
$\operatorname{REFL}_{i}-\left\langle\mathrm{REFL}_{i}\right\rangle$ integrate
'he became integrated'

An important property of this infixation pattern is that the infix does not undergo OCP(C). If a root starts with a sibilant and has a $/ \mathrm{t} / \mathrm{as}$ the second consonant, the reflexive form

[^24]on the surface seems to violate the $\operatorname{OCP}(\mathrm{C})$. Two non-final consonants inside the root are identical: the infixed part $\langle t\rangle$ of the reflexive marker and the second consonant $/ \mathrm{t} / \mathrm{of}$ the root. There is no repair strategy applied here to resolve this issue. The infixed $\langle\mathrm{t}\rangle$ does not become part of the domain of the $\operatorname{OCP}(\mathrm{C}) .{ }^{7}$
(13) Infix is not subject to $\mathrm{OCP}(\mathrm{C})$
a. hi-s $\langle t\rangle$ ater
$\mathrm{REFL}_{i}-\left\langle\mathrm{REFL}_{i}\right\rangle$ secret
'he hid himself'
b. hi- $\int\langle\mathrm{t}\rangle$ atef
$\operatorname{REFL}_{i}-\left\langle\mathrm{REFL}_{i}\right\rangle$-share
'he participated in s.th'

Crucially, this infix does not block the application of the $\operatorname{OCP}(\mathrm{C})$ on roots either. There are no forms where the two consonants separated by the infix can be identical. No alternation can be found where a root that would be neutralized with OCP satisfying roots in all other context shows up with the underlying consonants here. This pattern will be discussed in more detail in the next section.
(14) Non-existing infixed forms in Hebrew
a. ${ }^{* * h i-s}\langle\mathrm{t}\rangle$ asek
b. ${ }^{* *}$ hi- $\int\langle\mathrm{t}\rangle \mathrm{a}$ aek

The important generalization in Hebrew appears to be that only the root can ever be a domain for $\operatorname{OCP}(\mathrm{C})$, regardless of infixation. The infix does not become part of the domain and the discontiguous root stays a domain for the $\operatorname{OCP}(\mathrm{C})$ even if disrupted by an infix.

### 3.3 Constraints on Non-Contiguous Domains

In the preceding two sections, we have seen that Muna infixation and Hebrew Hitpa'el forms differ in one crucial aspect. Whereas in Muna repair strategies are applied in order to satisfy the *BM-constraint in the root-infix complex, Hebrew Hitpa'el forms apparently violate the $\mathrm{OCP}(\mathrm{C})$ constraint on the surface. In a more abstract way, this difference can be described in terms of domains. In Muna, the infix becomes part of the domain of the *BM-constraint, whereas in Hebrew the infix does not become part of the OCP $(\mathrm{C})$ domain.

[^25](15) Domains for root domain constraints in Hebrew and Muna infixed forms


There is another aspect, however, that Muna and Hebrew have in common. The root stays a domain if infixed. It does not lose it status as a domain. Discontiguous roots are still subject to the root domain constraint. Otherwise, we would expect that roots that underlyingly violate the root domain constraint can surface faithfully with this violating structure if infixed. In the non-infixed forms this violation would be repaired. An infixed form would thus seem to exhibit a phonological process that works against the root domain constraint by creating violating structures for a restricted set roots. Based on the surface representation in the output, this could be expected. Any locality relations that hold in the root are interrupted in the infixed form. If constraints make reference to these relations, we could expect these constraints to be relaxed. What would a concrete pattern look like?

Let us start by examining a hypothetical language Muna'. This language is similar to Muna in all and every aspect except that it relaxes the *BM-constraints for roots in infixed forms. Recall that in Muna, roots cannot have a labial obstruent as their first consonant if the second consonant is $/ \mathrm{m} /$. However, in order to examine what happens to such roots when the BM constraint is relaxed in infixed forms, we have to first establish what happens in non-infixed forms. For expository purposes, let us assume that root-internally violations of the *BM-constraint are repaired by dissimilating the place of articulation of $/ \mathrm{m} /$ from labial to a coronal $/ \mathrm{n} / .^{8}$ This means that a lexical form /fomi/ undergoes dissimilation and is mapped to surface form [foni]. The derivation for such forms is given in (16-a).

| Derivation of BM-violating lexical forms in Muna' |  |  |  |
| :--- | :---: | :---: | :---: |
| Input: | a. | /fomi/ | b. |
| /fomi/, $\langle\mathrm{um}\rangle$ |  |  |  |
| Infixation |  | - |  |
| BM-constraint: | foni |  | f-omi $\rangle$ omi |
| Output: |  | [foni] |  |

Now, what would happen if such forms undergo infixation? As mentioned before, the *BM constraint is loosened in infixed forms and does not apply to the domain consisting solely of the root anymore. This means that the domain of application of nasal accretion in (16-b) is now shifted to include the infix and the root-initial consonant. Interestingly, the $/ \mathrm{m} /$

[^26]that was banned from surfacing due to the BM constraint in the isolated root now surfaces. Whereas other roots in Muna' can undergo infixation just fine, roots that underlyingly violate the $* \mathrm{BM}$ constraint show up with an $/ \mathrm{m} /$ in addition to nasal accretion. Such alternating forms are not attested in Muna. They would be neutralized with other roots in any context but the infixation context. On the surface, it looks as if an anti-BM process applies across the infix. As far as I know, such a pattern is not attested in any natural language. In section 3.7, I provide a small scale study of 55 interaction patterns from 32 languages in the DoCoMD. I show that the Hebrew type and the Muna type interactions are robustly attested, whereas the Muna' type is not.

Why should this pattern be important at all? Why should we care about a hypothetical language that does not exist? I will show in section 3.5 that several theories predict such a pattern without further assumptions. This pattern will allow us to distinguish between different solutions. Some of these approaches allow for an easy way to get rid of this overgeneration, whereas others do not. Simply said, any procedural approach that allows ordering infixation before or simultaneous to the application of these constraints predicts a language Muna'. Procedural approaches can be restricted by the Strong Domain Hypothesis (Kiparsky 1985), but this makes wrong predictions for prefixes and suffixes, cf. section 3.5.1. Procedurally, infixes are not unusual, whereas representationally they are special in that they create non-contiguous domains. A fixed constraint ranking will be enough to exclude the Muna' pattern if Hierarchical Morphoprosodic Structure is assumed. A more detailed discussion will be provided in the next two sections.

The problem becomes more apparent if we look at the analogous counterpart for the Hebrew pattern: Hebrew'. This is the fourth logical possible language, where neither the discontiguous root, nor the root+infix complex form a domain for the $\operatorname{OCP}(\mathrm{C})$ in infixed forms. I mention this pattern here for the sake of completeness and to give another example of the effects we could observe when a root ceases to be a domain in infixed forms. Note that since Hebrew does not show any repair strategies, I will assume for the sake of exposition that Hebrew repairs $\operatorname{OCP}(\mathrm{C})$ violations by changing the mode of articulation of the violating consonant. Again, let us assume that Hebrew' has underlying forms that violate the $\mathrm{OCP}(\mathrm{C})$ but are repaired. In Hebrew proper they are repaired in every context, but in Hebrew' the repair strategy is suspended in case of infixation. This means that a hypothetical lexical form /sasek/ could be mapped to [sarek] in isolation, as shown in (17). If the reflexive form is constructed however, the infixation of $\langle t\rangle$ relaxes the $\mathrm{OCP}(\mathrm{C})$ and blocks the dissimilatory repair strategy. The resulting form is [histasek]. Note that the $/ \mathrm{s} /$, which was dissimilated in isolation, shows up in the infixed form. Judging from the alternating surface forms, it looks as if a an assimilation process takes place across the infix. Again, this hypothetical pattern is not attested among the languages of the world.

| Input: | a. /sasek/ | b. /hi- $\langle\mathrm{t}\rangle$ sasek/ |
| :---: | :---: | :---: |
| OCP(C) \& Infixation: | sarek | - |
| Output: | [sarek] | [histasek] |



In the last sections we have seen the four possible domain combinations for infixed forms. They vary according to two dimensions. The first potential difference concerns the inclusion of the infix into the domain of the constraint. This is the difference between Muna and Hebrew, and analogously between the two hypothetical languages Muna' and Hebrew'. Whereas the infix in Muna and Muna' undergoes a repair process, it remains unaffected in Hebrew and Hebrew ${ }^{\prime}$. Both options are attested. The other dimension reproduces the difference between attested forms and unattested forms. In both attested patterns Muna and Hebrew - the root remains a domain in infixed forms. A pattern that allows roots to cease being a domain is not attested. The effective root domains in infixed forms and the different interaction types are summarized in (19).
(19) Possible domain combinations in infixation

|  |  | Root+Infix Domain |  |
| :--- | :---: | :---: | :---: |
|  |  | Yes | No |
| Discontiguous Root domain | Yes | Muna | Hebrew |
|  | No | Muna' $^{\prime}$ | Hebrew $^{\prime}$ |

### 3.4 Hierarchical Morphoprosodic Structure

In this section, I will show that Hierarchical Morphoprosodic Structure is powerful enough to derive both the Muna pattern and the Hebrew pattern by different constraint rankings in a parallel OT system. ${ }^{9}$ This is accomplished by relativizing markedness constraint to different prosodic domains, defined by prosodic constituents and two kinds of domination relations. I will also show that Muna' can be excluded by a fixed ranking of these relativized markedness constraints.

Prosodic structure has been argued to explain domains of phonological processes (Inkelas 1990; Downing 1998; van Oostendorp 1999; Nespor \& Vogel 2012). The following approach

[^27]will be based on three assumptions: (i) Recursive Prosody: Recursive prosodic words exist in the output and can match roots, affixes, and words. It is assigned based on morphological affiliation of phonological material (van Oostendorp 2007; Selkirk 2011) (ii) Prosodic Domain Constraints: Constraints can be relativized to apply only inside a certain prosodic domain. (iii) Domination: Constraints can either apply to all material directly dominated by a constituent or indirectly dominated by a constituent

First of all, it needs to be shown that hierarchical structure allows reference to the root domain as well as the root+infix domain. I assume that an infixed form consists of a recursive structure, where the infix is dominated by a prosodic word ${ }^{10}$, which in turn is dominated by a prosodic word that also dominates the root. The smallest possible prosodic word that includes the root will be termed the prosodic root word $(\sqrt{\omega})$. This is a derived notion and no axiomatic primitive, see also chapter 1 for more discussion.
(20) An infixed form with Hierarchical Morphoprosodic Structure


This representation allows for two different domains if we allow for a distinction between direct domination and domination in general. Domination in general is able to ignore intervening prosodic categories of the same type, whereas direct domination does not apply across intervening categories. The domain defined by direct domination is thus a subdomain of the domain defined by domination in general. A definition of direct domination is given in (21). The domain defined by direct domination of $\sqrt{\omega}$ corresponds to the simple root. In the following, I will refer to this as the narrow domain of $\sqrt{\omega}$. Its wide domain, defined by domination in general, corresponds to the whole root+infix domain. Hierarchical prosodic structure thus allows reference to the two domains needed to describe the attested infixation patterns.

[^28]Direct Domination
Phonological material P is directly dominated by a prosodic category $\pi$ if
a. $\quad \pi$ dominates P and
b. there is no other prosodic category $\rho$ of the same type such that
(i) $\pi$ dominates $\rho$
(ii) $\rho$ dominates P

In order to allow for this representation to affect phonological computation in parallel Optimality Theory, constraints need to make reference to these domains. I will assume that any constraint potentially exists in two relativized versions for each domain, one for the wide domain (i.e. defined by general domination) and one for the narrow domain, defined by direct domination.

Relativized Markedness constraints
a. Markedness ${ }_{\pi, \text { DD }}$

Count one violation for each marked structure XY if X and Y are directly dominated by the same prosodic constituent $\pi$.
b. Markedness ${ }_{\pi, \mathrm{D}}$

Count one violation for each marked structure XY if X and Y are dominated by the same prosodic constituent $\pi$.

These constraint domains are enough to derive the Muna and the Hebrew pattern. For the Muna patterns, both versions of the *BM constraints relativized to prosodic root words have to be ranked above the corresponding faithfulness constraint. The ${ }^{*} \mathrm{BM}$ constraint thus applies both to the simple root and the root+infix domain. ${ }^{11}$ As shown in the tableau in (23), candidate ( $23-\mathrm{c}$ ) that only dissimilates the root consonants but does not change the infix, cannot become optimal due to a fatal violation of $* \mathrm{BM}_{\sqrt{\omega}, \mathrm{D}}$. The infix+root complex still violates the *BM constraint. Conversely, candidate (23-d) cannot become optimal, even though the infix has been deleted. The root consonants $/ \mathrm{f} /$ and $/ \mathrm{m} /$ trigger a markedness violation that keeps the structure from becoming optimal. Instead, candidate (23-b) where the infix has been deleted and the root consonants have been dissimilated can become optimal. The double violation of the low-ranked faithfulness constraint does not change this result. ${ }^{12}$

[^29](23)

| Input: fumomi | ${ }^{\mathrm{BM}_{\sqrt{\omega}, \mathrm{DD}}}$ | $*^{\mathrm{BM}_{\sqrt{\omega}, \mathrm{D}}}$ | FAITH |
| :---: | :---: | :---: | :---: |
|  | *! | * |  |
| 回 |  |  | ** |
|  |  | *! | * |
| d. | *! | *! | * |

The Hebrew pattern only needs a minimally different ranking. Since the constraint does not apply in the root+infix domain, the $\operatorname{OCP}(\mathrm{C})$ relativized to the wide domain of $\sqrt{\omega}$ needs to be ranked below the faithfulness constraint. As shown in (24), the faithful candidate (24-a) can become optimal, even though the infix causes a violation of the general constraint $\operatorname{OCP}(\mathrm{C})_{\sqrt{\omega}, \mathrm{D}}$. This is because the markedness constraint is ranked below FAITH, which is fatally violated by candidate (24-b), where a repair strategy has been applied.


| Input stater | $\mathrm{OCP}(\mathrm{C})_{\sqrt{\omega}, \mathrm{DD}}$ | FAITH | $\mathrm{OCP}(\mathrm{C})_{\sqrt{\omega}, \mathrm{D}}$ |
| :---: | :---: | :---: | :---: |
|  |  |  | * |
| b. |  | *! |  |

A third ranking is imaginable, where the wide domain constraint outranks the faithfulness constraint but the narrow domain constraint does not. This would predict the Muna' pattern, since the constraint would apply in a root+infix domain but not to the simple root. As shown in the tableau in (25), the same input /fumomi/ that yields [moni] in Muna, will result in [funomi] in Muna' under the ranking Markedness ${ }_{\pi, \mathrm{D}} \gg$ FAITH $\gg$ Markedness $_{\pi, \text { DD. }}{ }^{13}$ Candidate (25-b), which repairs ${ }^{*}$ BM violations in both domains, loses to candidate $(25-\mathrm{d})$ due to faithfulness violations. This is despite the $* \mathrm{BM}_{\sqrt{\omega}, \mathrm{DD}}$ violations of candidate ( $25-\mathrm{d}$ ) inside the discontiguous root.

[^30](25)

Muna' is predicted by Markedness ${ }_{\pi, \mathrm{D}} \gg$ FAIth $\gg$ Markedness $_{\pi, \mathrm{DD}}$

| Input: fumomi | * $\mathrm{BM}_{\sqrt{\omega}, \mathrm{D}}$ | FAITH | $*^{\text {BM }}{ }_{\sqrt{\omega}, \mathrm{DD}}$ |
| :---: | :---: | :---: | :---: |
|  | *! |  | * |
| b. |  | **! |  |
|  | *! | * |  |
|  |  | * | * |

Even though Hierarchical Morphoprosodic Structure thus faces a problem similar to the other approaches, this is not fatal. The Muna' pattern can be easily excluded by a fixed ranking of relativized markedness constraints in general, $M_{D D} \gg M_{D}$. This fixed ranking has clear empirical effects. If a structure is forbidden in a wide domain of a prosodic constituent it is also forbidden in its narrow domain. This can be linked to other fixed rankings that have been proposed to explain locality effects. Suzuki (1998) for example proposes a fixed ranking, where more local OCP constraints always outrank less local OCP constraints, see also Stanton (2020) for a recent reanalysis of counterexamples.

Note that such a fixed ranking does not exclude derived environment effects per se. Its predictions are thus different from the Strong Domain Hypothesis (Kiparsky 1985), cf. section 3.5.1. The reason is that prosodic constituents are recursive and prefixes and suffixes can be included in a higher domain of the same type, cf. the structure in (26-b). Instead of deriving the pattern by differences in constraints, the difference would thus be structural. ${ }^{14}$ Note that this is impossible for infixes. They are representationally special

[^31]because they are included inside a domain and do not occur at its edge. Including them only in a higher domain would lead to a prosodic word included inside a prosodic word that is not its mother, i.e. a structure akin to crossing association lines. This is illustrated in (26-a).
(26) Infixes and other affixes included in higher prosodic constituents
a. Infixes cannot be exclusively included in higher domains

b. Other affixes can be exclusively included in higher domains


A final restrictions that needs to in place concerns the possible representations. If a hierarchical structure such as the one in (27) were allowed, it would generate the Muna' pattern independent of the type of domination included. In order to explain the absence of this prosodic structure, we need to formalized the mechanism that generates morphoprosodic structures. Based on van Oostendorp (2007), I assume that morphoprosodic structure is generated on the basis of morphological color, i.e. morphological affiliation, which is already present in the input representation. I propose that the illicit structure can be excluded, since it is harmonically bound by the structure discussed above in most contexts. The reason is that the illicit structure violates the $\operatorname{MATCH}(\alpha, \pi)$ constraint that requires each morpheme to form its own prosodic word (van Oostendorp 1999; Selkirk 2011), whereas the correct structure does not. ${ }^{15}$ The $\operatorname{MATCH}(\alpha, \pi)$ requires each morpheme to be matched by a prosodic constituent, the reverse constraint $\operatorname{Match}(\pi, \alpha)$ requires each prosodic constituent to match a morpheme. The relevant structures are given in (27). ${ }^{16}$

[^32](27) Match constraints favor recursive structure for infixes
a. Ilicit structure: $\operatorname{Match}(\alpha, \pi) \boldsymbol{x}, \operatorname{Match}(\pi, \alpha)$

b. Recursive structure: $\operatorname{Match}(\alpha, \pi) \boldsymbol{V}, \operatorname{Match}(\pi, \alpha)$


The non-recursive structure thus violates an additional constraint. Nevertheless, it must also be guaranteed that there is no constraint that independently favors the non-recursive structure. This means that anti-recursivity constraints cannot be part of the constraint system as well as *STRUCTURE constraint specific to prosodic words. Gouskova (2003) has independently argued that *Structure constraint need to be absent from CON, since they predict unattested patterns. Constraints against recursive prosodic domains might not be needed at all, since in the case of prefixes and suffixes, match theory's Match constraints are already violated by recursive structure. Here we return to infixes being special representationally. Whereas recursive structure for an infix allows the edges of both morphemes to be matched by a recursive prosodic word (cf. (28-a)), this is impossible for prefixes and suffixes. In the latter case, a recursive structure will always violate some match constraint, since the higher prosodic word only matches maximally one edge of the root, as shown in (28-b,c).

[^33](28) Satisfaction of MATCH-constraints in recursive structure with infixes and prefixes
a. Infixes: $\operatorname{Match}(\alpha, \pi) \boldsymbol{V}, \operatorname{Match}(\pi, \alpha)$

b. Prefixes: $\operatorname{Match}(\alpha, \pi) \boldsymbol{\aleph}, \operatorname{Match}(\pi, \alpha)$

c. Prefixes: $\operatorname{Match}(\alpha, \pi) \boldsymbol{V}, \operatorname{Match}(\pi, \alpha)$


In sum, the representational approach can be easily restricted in order to differentiate between infixes and other affixes. The reason is that infixes are intrinsically special in representational terms because they form a smaller domain inside a larger domain without being at its edge. Therefore, an approach based on Hierarchical Morphoprosodic Structure only needs a fixed ranking to derive the correct empirical picture. In the next section, I will describe the problems that occur with procedural approaches to root domain constraints, followed by an evaluation of more restricted representational solutions in section 3.6.

### 3.5 Overgeneration of Procedural Approaches

Procedural approaches generally overgenerate with respect to the interaction between infixation and root domain constraints. SPE-like rule based phonology, Stratal Optimality and Harmonic Serialism predict the Muna' pattern. The reason is that they allow infixation to happen before a process that applies to the infix and the root. Apart from that, most procedural frameworks are a good fit for the data. SPE and Stratal OT can easily generate Muna and Hebrew, since they allow for non-surface true generalization encoded by opaque interactions. Neither of the two theories derive the unattested Hebrew' pattern. Both can in principle be restricted by the Strong Domain Hypothesis, which correctly excludes
the Muna' pattern but has been proven to make the wrong empirical predictions in other contexts for prefixes and suffixes. Again, this shows that infixes cannot be easily distinguished based on procedural terms, whereas representationally they are intrinsically different.

Harmonic Serialism (HS) is different from Stratal OT in that it does not generally allow opaque rule interactions. Therefore it faces difficulties in deriving the Hebrew pattern. Similarly, Optimality Theory with Output-Output Correspondence (OT-OO) faces additional challenges, since it fails to derive the simple Muna pattern unless equipped with additional representational power. Again, it has no means to derive the Hebrew pattern.

Predictions of procedural approaches

|  | SPE | Stratal OT | HS | OT-OO |
| :--- | :--- | :--- | :--- | :--- |
| Muna | $\boldsymbol{\nu}$ | $\boldsymbol{\nu}$ | $\boldsymbol{\nu}$ | $\boldsymbol{\nu}$ |
| Hebrew | $\boldsymbol{\nu}$ | $\boldsymbol{\nu}$ | $\boldsymbol{x}$ | $\boldsymbol{x}$ |
| Muna $^{\prime}$ | $\boldsymbol{\nu}$ | $\boldsymbol{\nu}$ | $\boldsymbol{\nu}$ | $\boldsymbol{x}$ |
| Hebrew $^{\prime}$ | $\boldsymbol{x}$ | $\boldsymbol{x}$ | $\boldsymbol{x}$ | $\boldsymbol{x}$ |

In this section, I will lay out the details of the overgeneration argument against procedural approaches. On the one hand, I will show for each approach which patterns it generates and why. On the other hand, I will show what patterns cannot be generated and present a principled account for the reason. All in all, I will investigate the exact predictions and flaws of each approach and explain why they cannot be easily overcome.

### 3.5.1 SPE with Morpheme Structure Constraints

Phonological approaches in the tradition of Halle (1959) and Chomsky \& Halle (1968) make a distinction between rules that restrict the underlying forms (Morpheme Structure Rules) and rules that map the underlying forms onto the surface forms (P-rules). Morpheme Structure Rules (MSR) apply to monomorphemic domains before morphological concatenation. P-rules generally apply after morphological concatenation. SPE now predicts three different languages from three different grammars, depending on the status of the rules that apply in Muna and Hebrew. They could be either MS-rules, P-rules, or both. Each of these predicted languages maps directly to one of the languages discussed in section 3.3. Only Hebrew' is not predicted by any rule order. ${ }^{17}$

[^34]| Predicted languages by rule orders in SPE |  |  |
| :--- | :--- | :--- |
| language | rule order | Interaction |
| Muna | MSR $\prec$ Infixation $\prec \mathrm{PR}$ | Feeding and Counterbleeding |
| Hebrew | MSR $\prec$ Infixation | Counterfeeding and Counterbleeding |
| Muna $^{\prime}$ | Infixation $\prec$ PR | Feeding and Bleeding |
| Hebrew $^{\prime}$ | - | - |

The Muna pattern can be easily predicted by an SPE-like approach, even though it involves the most complex derivations. The rule repairing the *BM-constraint (BM-rule from now on) has to be a MS-rule as well as a P-rule. This means that it applies before and after infixation. Therefore infixation creates a context for application of the BM-rule, i.e. infixation feeds the BM-rule. At the same time, however, the BM-rule is counterbled by infixation. If infixation would apply only before the BM-rule, infixation would destroy the root-internal context for application of the BM-rule by intervening between target and context.

| Derivation of Muna in SPE with MSC |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Lexical Form | dadi | foni | fomi | fomi |
| MSR | - | - | foni | foni |
| Infixation | $d\langle u m\rangle$ adi | $\mathrm{f}\langle u m\rangle$ oni | - | $\mathrm{f}\langle u m\rangle$ oni |
| P-rule | - | m -oni | foni | m -oni |
| Output | [dumadi] | $[$ moni] | [foni] | $[$ moni] |

Recall that in Hebrew the infix does not become part of the domain of the OCP(C). This can be modeled as a counterfeeding and counterbleeding interaction. The OCP(C)rule only applies to underlying forms before infixation. After infixation, this rule does not apply as a P-rule. Two interactions follow. First, infixation applies too late to provide a new context for the $\mathrm{OCP}(\mathrm{C})$-rule and thus the infix is not affected. Infixation counterfeeds the $\operatorname{OCP}(\mathrm{C})$-rule. Additionally, just like in the derivation of Muna, infixation also counterbleeds the $\mathrm{OCP}(\mathrm{C})$-rule. If the order were reversed, infixation would block application of the $\operatorname{OCP}(\mathrm{C})$-rule to the root domain. Hebrew can thus be derived as a doubly opaque interaction between Infixation and the $\operatorname{OCP}(\mathrm{C})$-rule.

| Derivation of Hebrew in SPE with MSC |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Lexical Form | sarek | sater | sasek | sasek |
| MSR | - | - | sarek | sarek |
| Infixation | hi-s $\langle t\rangle$ arek | hi-s $\langle t\rangle$ ater | - | hi-s $\langle t\rangle$ arek |
| Output | $[$ histarek $]$ | $[$ histater $]$ | $[$ sarek $]$ | $[$ histarek $]$ |

The crucial overgeneration problem lies in the prediction of the Muna' pattern. How does an SPE-like theory derive the Muna' pattern? Assuming the only rule order left, namely a P-rule applying after infixation generates this pattern. This results in a transparent feeding and bleeding pattern. Infixation provides the context for the BM-rule to apply to the root+infix complex. At the same time, infixation destroys the context for such a rule to apply to the root alone. The BM-rule applies in such a way to provide a surface-true generalization. This is a problematic result, since the Muna' pattern is not attested in the languages of the world, cf. section 3.7. The discontiguous root has ceased to be a domain in this root-infix-complex. The root domain constraint does not apply to discontiguous domains.

| Derivation of Muna' in SPE with MSC |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Lexical Form | dadi | foni | fomi | fomi |
| Infixation | $\mathrm{d}\langle u m\rangle$ adi | $\mathrm{f}\langle\mathrm{um}\rangle$ omi | - | $\mathrm{f}\langle\mathrm{um}\rangle$ omi |
| P-rule | - | m-oni | foni | m -omi |
| Output | [dumadi] | $[$ moni] | [foni] | $[$ momi] |

There is a simple way to exclude Muna' from the space of predictions that has been discussed before in the light of other empirical challenges. The only statement needed is a ban on rules applying at a later level without applying at an earlier level. Kiparsky (1985) already expressed the idea that rules might only be 'switched off' in the course of a derivation, never 'switched on'. This later became known as the Strong Domain Hypothesis, cf. (34).

## Strong Domain Hypothesis (cf. Kiparsky 1985:87)

It is possible to restrict the marking of domains to specifications of the form 'rule R does not apply after level n'.

In our case such a statement would correctly rule out Muna', since in Muna' the BM-rule only aplies after infixation but not before it. Unfortunately, the Strong Domain Hypothesis has been shown to be empirically untenable for data from other areas of phonology in many languages. For example, syllabification in Malayalam (Dravidian, India) has been argued to include closed syllables only at earlier levels, whereas on later levels all syllables are open (Mohanan 1989). This means that [ə]-epenthesis in order to create open syllables only takes place at a later level but not at an earlier level. The evidence comes from stress patterns as well as $\left[\mathrm{r}^{\mathrm{j}}\right] \sim[\pi]$ alternation and syllable structure restrictions, all of which must make crucial reference to the coda position in a syllable. [ə]-epenthesis thus cannot apply at an earlier level before affixation. Instead it must be 'switched on' at a later
stage after affixation has applied, thereby violating the strong domain hypothesis. Further evidence against the strong domain hypothesis comes from Turkish Velar Drop (Inkelas \& Orgun 1995), Dagbani High Tone Spreading (Hyman 1993), s-Deletion in Kaska and Sekani (Kaisse 1993), vowel raising in Baque (Hualde 1989) and many others. In order to correctly restrict a theory such as SPE, we would thus have to relativize the Strong Domain Hypothesis to infixation.

Revised Strong Domain Hypothesis (relativized to infixation)
Rules that apply after infixation also have to apply before infixation.

It should be clear from the formulation of this hypothesis that it is an ad-hoc device to capture the data. ${ }^{18}$ It restates the facts, i.e. that Muna' - a language where a rule applies after infixation but not before it - should be excluded. It makes no further predictions beyond the scope of interaction of phonological rules with infixation. Such an ad-hoc stipulatory restriction is undesirable.

### 3.5.2 Stratal Optimality Theory

Stratal Optimality Theory suffers from the same problem that hinders an SPE approach, only even more so. In addition to allowing different sequential orderings of infixation, it allows infixation and a repair process to apply simultaneously at the same stratum. This always yields a fully transparent pattern, i.e. Muna'. Needless to say, this is highly problematic. The tableau in (36), shows that simply ranking the *BM-constraint above a faithfulness constraint will always favor candidate (36-b), where the constraint only applies in the root+infix domain, over the attested candidate (36-d), where dissimilation has also applied in the root. The additional faithfulness violation of candidate (36-d) is not motivated by any higher ranked constraint and turns out to be fatal.

Monostratal Transparency generates the Muna' pattern

| I: | $\mathrm{f}\langle u m\rangle$ omi | *BM | FAITHFULNESS |
| ---: | :--- | :---: | :---: |
| a. | fumomi | $*!$ |  |
| b. | m-omi |  | $*$ |
| c. | fumoni | $*!$ |  |
| d. | m-oni |  | $* *!$ |

The same ranking also predicts that the *BM constraint is repaired in simple isolated

[^35]roots. In the tableau in (37), the faithful candidate (37-a) violates the *BM constraint. Therefore, the candidate (37-b) that repairs the constraint violation becomes optimal.

Monostratal Transparency generates the Muna' pattern

| I: | fomi | *BM | FAITHFULNESS |
| ---: | :---: | :---: | :---: |
| a. | fomi | $*!$ |  |
| b. foni |  | $*$ |  |

Additionally, all serial interaction employed by an SPE-like grammar can be transferred to a Stratal OT account. Any stratum with a ranking of markedness constraint over faithfulness constraints $(M \gg F)$ allows for a repair process to apply whereas any stratum with the reverse ranking $(F \gg M)$ blocks such a process. If a stratum with $M \gg F$ succeeds infixation at a stratum with the ranking $\mathrm{F} \gg \mathrm{M}$, we generate $\mathrm{Muna}^{\prime}$ as a feeding pattern. The *BM-constraint does not need to be satisfied when infixation applies. Later on, the whole infix+root complex is evaluated and needs to satisfy the *BM-constraint. The root is thus not a domain if infixed. The predictions are summarized in (38). ${ }^{19}$

Patterns predicted by rankings in Stratal OT
a. Infixation at Stratum 1

Stratum 1

|  |  | $M \gg F$ | $\mathrm{~F} \gg \mathrm{M}$ |
| :--- | :--- | :--- | :--- |
| Stratum 2 | $\mathrm{M} \gg \mathrm{F}$ | Muna $^{\prime}$ | $\mathrm{Muna}^{\prime}$ |
|  | $\mathrm{F} \gg \mathrm{M}$ | Muna | - |

b. Infixation at Stratum 2

$$
\text { Stratum } 1
$$

|  |  | $\mathrm{M} \gg \mathrm{F}$ | $\mathrm{F} \gg \mathrm{M}$ |
| :--- | :--- | :--- | :--- |
| Stratum 2 | $\mathrm{M} \gg \mathrm{F}$ | Muna | $\mathrm{Muna}^{\prime}$ |
|  | $\mathrm{F} \gg \mathrm{M}$ | Hebrew | - |

Going through the concrete implementations of Muna' separately, we could assume that infixation applies at Stratum 1, where the *BM constraint is ranked above the faithfulness constraint. This will derive the Muna' pattern by parallel interaction between infixation and root domain constraints, as shown in the tableaux in (36) and (37). A repair mechanism will apply in parallel with infixation. This generated pattern is independent of the ranking at the second stratum.

[^36]Another possible implementation of Muna' consists of applying infixation at the first stratum, where the faithfulness constraint is ranked above the markedness constraint. No repair mechanism applies at Stratum 1. At Stratum 2, the ranking is reversed. The markedness constraint is now ranked above the faithfulness constraint. Since infixation has already applied, the whole root+infix domain undergoes the repair mechanism. The discontiguous root on the other hand is not subject to the root domain constraint anymore. The Muna' pattern is therefore derived as an interstratal feeding relation.
(39) Derivation table for Muna' by interstratal feeding interaction

| Input | $/$ fomi/,/um/ | /fomi/ |
| :--- | :--- | :--- |
| Stratum 1: $\mathrm{F} \gg \mathrm{M}$, Infixation | $\mathrm{f}\langle\mathrm{um}\rangle$ omi | fomi |
| Stratum 2: M $>\mathrm{F}$ | m -omi | foni |
| Output | [momi] | [foni] |

A final possible implementation assumes that both infixation and the repair mechanism apply at Stratum 2. This again yields the parallel interaction from the tableau in (36). Infixation applies at a stratum where the markedness constraint is ranked above the faithfulness constraint. Therefore, the repair mechanism applies to the whole root+infix domain and not to the discontiguous root.

Note that restricting Stratal OT is much more difficult. A version of the Strong Domain Hypothesis, which has been proposed in work on Stratal OT, restricts reranking between strata (cf. Itô \& Mester 1999). ${ }^{20}$ Since promoting a faithfulness constraint above a markedness constraint will block a process at the later stratum, this process is parallel to the 'switching off' of rules in rule-based phonology. ${ }^{21}$ Note however that the restriction in (40) is not enough to exclude the Muna' pattern.

## Strong Domain Hypothesis for Stratal OT

Reranking between strata only consists of promoting faithfulness constraints above markedness constraints.

The problematic cases are the ones where simultaneity of infixation and the ranking $\mathrm{M} \gg \mathrm{F}$ generates a Muna' pattern by monostratal transparency. Muna' is not only derived by

[^37]serial interaction but also by applying infixation and a repair mechanism for the root domain constraint at the same stratum. These interactions are crucially not excluded by the restriction in (40), since it only talks about interstratal reranking. Stratal OT thus has a serious overgeneration problem.

### 3.5.3 Harmonic Serialism

The main additional problem of Harmonic Serialism is its inability to derive the opaque Hebrew pattern. As has been noted, Harmonic Serialism strongly favors transparent over opaque derivations. More specifically, Harmonic Serialism cannot derive a counterfeeding pattern (McCarthy 2000), like the one we see in Hebrew. The argument runs as follows. If the $\mathrm{OCP}(\mathrm{C})$ constraint is ranked higher than the faithfulness constraint when evaluating the root on itself, the first step will involve a repair of an $\mathrm{OCP}(\mathrm{C})$-violating root, as shown in (41). Nevertheless, after infixation has applied, the OCP(C) constraint will need to be satisfied again, since the ranking has not changed, cf. (42). The infix thus cannot be excluded from the domain of the $\mathrm{OCP}(\mathrm{C})$ constraint. This makes it impossible to derive the Hebrew pattern. ${ }^{22}$
Step 1 for a derivation of /sasek/

| I: | /sasek/ | OCP $(\mathrm{C})$ | FAITHFULNESS |
| ---: | :--- | :---: | :---: |
| a. | sasek | $*!$ |  |
| b. | sarek |  | $*$ |

Step 3 (after infixation) for a derivation of /sater/

| I: | /histater $/$ | OCP $(\mathrm{C})$ |
| ---: | :---: | :---: |
| FAITHFULNESS |  |  |
| a. | histater | $*!$ |
| -b. | histarer |  |

In order to derive the Hebrew pattern, the theory would have to be considerably expanded. One possibility would be to introduce constraints indexed to affixes or roots, a distinction that would have to be coded in the representations, e.g. by indexing of phonological material that belongs to a root as $\mathrm{P}_{\mathrm{R}}$. $\mathrm{An} \mathrm{OCP}(\mathrm{C})_{\mathrm{R}}$ constraint indexed to roots would then not be violated by an intervening infix. Note that these representations and constraints are not readily compatible with an analysis of the Muna pattern, as I show in section 3.6.

As for the other patterns analyzed in Harmonic Serialism, it is important to note that there is one striking point of departure between different analyses in Harmonic Serialism that involve morphologically complex forms. Some analyses (implicitly) assume an account

[^38]where all phonological evaluation follows all morphological operations (Elfner 2016; TorresTamarit 2016) whereas others assume an approach where morphology and phonology are interleaved (cf. e.g. McCarthy 2012; Gleim et al. 2022). If interleaving is impossible, only the Muna' pattern is predicted because infixation will necessarily precede any repair mechanism triggered by the *BM-constraint. Interleaving does not rule out Muna' because infixation can still apply before the first application of phonology. The Muna' pattern is thus again predicted by a simple feeding relationship between infixation and a repair mechanism for root domain constraints. In sum, Harmonic Serialism undergenerates the Hebrew pattern and overgenerates the Muna' pattern.

### 3.5.4 Output-Output-Correspondence

OT with Output-Output correspondence (Benua 1997; McCarthy 1998) allows for overapplication in derived domains via a faithfulness constraint to roots. It can thus generate a pattern similar to the Muna pattern. In roots, the *BM-constraint is satisfied transparently by ranking the *BM constraint above the faithfulness constraint. Therefore /fomi/ is mapped to [foni].

In the evaluation of the infixed form (cf. (42)), a new faithfulness constraint becomes active that protects the root by relating it to its base, i.e. [foni], a non-infixed output where *BM has been satisfied. Therefore, the domain of the *BM-constraint is non-contiguous in infixed forms. The constraint additionally applies to the infix, since *BM is still ranked above the general faithfulness constraint.

Output-Output Faithfulness generates a pattern similar to Muna

| I: f(um)omi <br> Base: foni | *BM | OO-Faithfulness | IO-FAIthfulness |
| :---: | :---: | :---: | :---: |
| a. fumomi | *! | * |  |
| b. moni |  |  | ** |
| c. fumoni | *! | * | * |
| d. momi |  | *! | * |

Additional constraints are needed to keep other affixes from undergoing the constraint. If such constraints where not added, a different pattern would emerge. I will call this the Sundanese pattern, since it is attested in Sundanese (Austronesian, Indonesia) nasal harmony (Cohn 1992; Benua 1997). A process applies everywhere transparently and overapplies in a root across an an infix. The Sundanese pattern is different from the Muna pattern where a process applies only to roots (independently of infixation) and the root+infix complex but not in other contexts.

| Possible Rankings and predicted languages in OT-OO |  |
| :--- | :--- |
| Ranking | Predicted Pattern |
| OO-FAITH $\gg$ BM $\gg$ IO-FAITH | Sundanese |
| *BM $>$ OO-FAITH $>$ IO-FAITH | Sundanese |
| OO-FAITH $>$ IO-FAITH $\gg$ BM | $*$ BM inactive |
| IO-FAITH $\gg$ IO-FAITH $\gg$ BM | $*$ BM inactive |
| IO-FAITH $\gg$ BM $>$ OO-FAITH | $*$ BM inactive |
| $*$ BM $>$ IO-FAITH $>$ OO-FAITH | general *BM-constraint |

No further pattern can be derived since we can construct the factorial typology in (44), based only on these three constraints. If the markedness constraint and OO-faithfulness are ranked above IO-faithfulness, the Sundanese pattern is generated. If the general faithfulness constraint is ranked above the markedness constraint, the process will be blocked across the board. If *BM is ranked above IO-faithfulness, which in turn outranks OO-faithfulness, the repair will apply in all contexts, regardless of any properties of the base. All in all, OT with Output-Output correspondence is a very restrictive theory but its predictions do not match up with the empirical picture.

Similar to the Harmonic Serialism case, the Hebrew pattern can only be derived if the theory is expanded. As hinted at above, a further faithfulness constraint is needed for independent reasons. If phonological material is indexed to its root and affix status, an affix specific faithfulness constraint is possible (Ussishkin 2005; Hargus \& Beavert 2006). The representational power that is added allows us to derive the Hebrew pattern. In the tableau in (45), both the root and the affix are protected in the infixed form by designated faithfulness constraints. The $\mathrm{OCP}(\mathrm{C})$ can thus not be repaired, neither by changing the root in candidate (45-b) nor by changing the affix in candidate (45-c). The OCP can still apply in non-infixed roots.

Affix-Faithfulness in OT-OO

| I: / $\mathrm{h}_{\mathrm{A}} \mathrm{i}_{\mathrm{A}} \mathrm{St}_{\mathrm{A}}$ ater/ <br> Base: sater | $\mathrm{FAITH}_{\text {Affix }}$ | Faith-OO | $\mathrm{OCP}(\mathrm{C})$ | Faith-IO |
| :---: | :---: | :---: | :---: | :---: |
| ara. $\mathrm{h}_{\mathrm{A}} \mathrm{i}_{\mathrm{A}} \mathrm{St}_{\mathrm{A}}$ ater |  |  | * |  |
| b. $\mathrm{h}_{\mathrm{A}} \mathrm{i}_{\mathrm{A}} \mathrm{st} \mathrm{A}_{\mathrm{A}}$ arer |  | *! |  | * |
| c. $\mathrm{h}_{\mathrm{A}} \mathrm{i}_{\mathrm{A}} \mathrm{Sr}_{\mathrm{A}}$ ater | *! |  |  | * |

However, the same constraint leads to a ranking paradox if one attempts to derive the Muna pattern. The markedness constraint *BM needs to outrank affix faithfulness because infixes undergo a repair mechanism to satisfy it. *BM must also be ranked higher than the general faithfulness constraint Faith-IO in order to apply to isolated roots and infixes.

However, affixes other than the infix need to be protected by ranking at least one of these faithfulness constraints above the *BM constraint. One possible solution to this problem is a further splitting of faithfulness constraints into a faithfulness constraint indexed to the infix $\langle u m\rangle$ and a faithfulness constraint indexed to all other affixes, with the latter being ranked above the *BM markedness constraint. This, however, misses the generalization that in Muna only infixes can be affected by root domain constraints. Furthermore, modularity is weakened by direct reference to a specific morpheme.

### 3.6 Undergeneration of Representational Solutions

In the following section, I will present different representational devices that have been used in order to determine domains in phonology. I will show that boundary markers are incapable of deriving the interaction of infixation and root domain constraints in a satisfactory way. Such an approach predicts unattested patterns and fails to predict attested patterns. Even though morphological colors alone only predict the Hebrew pattern, they are a necessary ingredient for more complex representations. For the sake of comparison, I will assume that these representational solutions are combined with Standard Parallel Optimality Theory and allow for a formulation of novel constraints that make reference to these representations.

| Predictions of representational approaches |  |  |
| :--- | :--- | :--- |
|  | Boundary Markers | Morphological Colors |
| Muna | $\boldsymbol{x}$ | $\boldsymbol{*}$ |
| Hebrew | $\boldsymbol{x}$ | $\boldsymbol{\checkmark}$ |
| Muna $^{\prime}$ | $\boldsymbol{\swarrow}$ | $\boldsymbol{\aleph}$ |
| Hebrew $^{\prime}$ | $\boldsymbol{\swarrow}$ | $\boldsymbol{\aleph}$ |

### 3.6.1 Boundary Markers

Boundary markers have been employed in SPE (Chomsky \& Halle 1968) to restrict the application of rules to certain domains. They consist of a symbol that is linearized in a string with all other segments. As first noted by McCarthy (1979:227), boundary markers make wrong predictions with regard to infixation and root domain constraints. An infix splits up the root domain because it needs to be enclosed in boundary symbols. Since there is no other information present in the string, any constraint making reference to these boundary markers will be relativized to a domain defined by these boundary markers, i.e. a string enclosed by it. In the general case, a discontiguous root is not a possible domain here.

Take the formulation of the markedness constraint in (47). When applied to an infixed
string such as the example in (48), it predicts three contiguous domains for the application of the constraint, the first root consonant $\# \mathrm{~s} \#$, the infixed $\# \mathrm{t} \#$ and the rest of the root \#ater\#. The root is not a domain anymore if infixed. Discontiguous roots are not domains for root domain constraints. This means that the constraint does not apply inside infixed roots at all. The domains are too small. We thus predict a Hebrew' pattern, where the constraint applies to uninfixed roots but not to infixed roots.

## Markedness Constraint $\# \#$

Count one violation for each marked structure XY that is located between two boundary markers.

$$
\begin{align*}
& \text { \#hi\#s } \# \mathrm{t} \# \text { ater } \#  \tag{48}\\
& \left\langle\mathrm{REFL}_{i}\right\rangle \text { secret } \\
& \text { 'he hid himself' }
\end{align*}
$$

The boundary marker theory can be extended to include specific root boundary markers ' + ', which only occur to the left and to the right of roots. This means that the constraint can be reformulated to only apply between root boundary markers, as in (49). The same example now is interpreted differently by the constraint. As shown in (50), we can now identify the root domain very easily. The root domain includes the segments /stater/ stemming both from the infix and the discontiguous root. If the constraint transparently applies in this domain, however, the result is a Muna' pattern (cf. section 3.5.2). The reason is that the discontiguous root is not a domain for the constraint. The constraint only applies in the root+infix domain. ${ }^{23}$

## Markedness Constraint ${ }_{++}$

Count one violation for each marked structure XY that is located between two root boundary markers.
\#hi+s\#t\#ater+

It would thus be necessary to explicitly exclude the infix from the domain of the constraint. This would not only be an ad-hoc complication of the constraint, the constraint itself would only derive the Hebrew pattern. As formulated in (51), the domain of the constraint would only include the simple root /sater/ and ignore the infix altogether. In order to derive the Muna pattern, both versions, (49) and (51), would need to be included.

[^39]
#### Abstract

Markedness Constraint $+\ldots+$ Count one violation for each marked structure XY that is located between two root boundary markers and that is not seperated by or located between two regular boundary markers.


I concur that the formulation of the constraint in (51) is an ad-hoc complication that only serves to derive the data on interaction between root domain constraints and infixation. It does not add any explanatory force.

### 3.6.2 Morphological Colors

Morphological colors (van Oostendorp 2007; Revithiadou 2007) are part of the larger class of morphologically indexed representations, where phonological objects are indexed for their morphological affiliations. Approaches that need similar representations are Rootand Affix-Faithfulness (McCarthy \& Prince 1995) as well as lexical constraint indexation (Pater 2007). Morphological colors are a specific version of such a theory because it is assumed that phonology cannot manipulate morphological affiliation as represented by morphological affiliation. Additionally, the assumption is that reference to morphological information is minimal in phonology, i.e. morphological colors cannot be fully interpreted in phonology. Phonology can distinguish between different morphemes but it cannot refer to specific morphological information. This is in stark contrast to approaches like lexical constraint indexation (Pater 2007) where constraints can refer to individual morphemes. In (52), the Hebrew data are converted into a representation with morphological colors. The root is represented in black, whereas the reflexive infix is given in blue.
histater
. $\mathrm{REFL}_{i}\left\langle\mathrm{REFL}_{i}\right\rangle \mathrm{secret}^{\prime}$
'he hid himself'

Such a representation allows us to formulate a markedness constraint that applies to the root domain in a simple manner. We only have to restrict it to applying only to phonological objects of the same color, as in (53). Such a constraint gives us two domains, namely the infix and the root. Both domains are separated and the constraint does not apply to the root+infix domain. This is enough to derive the Hebrew pattern.
(53) Markedness Constraint ${ }_{\text {MC }}$

Count one violation for each marked structure XY if both X and Y are of the same morphological color.

Any attempt to derive the Muna pattern would entail serious complications. A domain that includes both the root and the infix would need to be defined by making reference to the first and last segment in a string with a given morphological color. Note the similarity to the complex constraint for boundary markers. Again, I note that the introduction of such a constraint type is not motivated by any data outside of infixation and root domain constraints.


#### Abstract

Markedness Constraint ${ }_{\text {M...C }}$ Count one violation for each marked structure XY if both X and Y occur inside a domain that starts with the leftmost segment A and ends with the final segment B, where both A and B are of the same morphological color.


Morphological Colors are thus a very restrictive representational device for modeling the interaction between infixation and root domain constraints. There is a serious undergeneration problem, since it does not derive the Muna pattern without any further additions.

### 3.7 Crosslinguistic Data

In this section, I will present further data on the interaction between root domain constraints and infixation. This serves two purposes. For one, it strengthens the claim that Hebrew ${ }^{\prime}$ and Muna' patterns are absent. Additionally, I will show that even though the language family has a strong effect on the pattern, Austronesian languages as well as Semitic languages show variable patterns. I will also show that root domain constraints interact with infixation in languages outside of these two language families.

In the DoCoMD database, the sources for 53 entries include information of the interaction of a constraint on monomorphemic domains with infixation. ${ }^{24}$ The entries include 32 languages from nine language families. The full infixation sample is provided in the appendix of this thesis. I included a pattern of interaction if the following conditions were met. The infixation pattern and the root domain constraint were explicitly mentioned in the source at hand. Forms with potentially violating infixes where given or mentioned. If these forms were attested, I concluded that there is no repair mechanism. If there was an explicit mention of a repair mechanism, I included these. This allows for a distinction between the Muna and the Hebrew pattern. If infixes did not have the potential to violate the constraints, I listed them as non-interacting. Similarly, I looked at alternations found in roots that were conditioned by infixation. The Hebrew' type and the Muna' type required

[^40]root alternations that would produce results that violate the root domain constraints in the given language. If I had found such a pattern, I would have classified it accordingly.

As shown in (55), most infixes do not interact with constraints on monomorphemic domains ( $\mathrm{n}=23$ ). Of the interacting infixes, the Muna-type is most common type ( $\mathrm{n}=20$ ), but the Hebrew-type is not uncommon ( $\mathrm{n}=12$ ). Crucially, the Muna'-type and the Hebrew'-type are unattested.

| Interaction between CoMDs and infixation in DoCoMD |  |
| :--- | ---: |
| Interaction Type | $\#$ of patterns |
| No Interaction | 23 |
| Muna-type | 20 |
| Hebrew-type | 12 |
| Muna'-type | 0 |
| Hebrew'-type | 0 |
| Total | 55 |

The areal distribution mainly mirrors the generally areally skewed distribution of infixation (cf. Yu 2007; Blevins 2014). Infixation is more common in Papuanesia but rarely described in Australia and South America. There are no entries with infixation from the latter two macro-areas and the majority of infixing languages in DoCoMD is from Papuanesia ( $\mathrm{n}=22$ ), cf. (56).
(56) Areal distribution of infixation in DoCoMD

| Macro-Area | \# of languages |
| :--- | ---: |
| Papuanesia | 22 |

Eurasia 7
North America 2
Africa 1

| Australia | 0 |
| :--- | ---: |
| South America | 0 |
| Total | 32 |

The different types of interaction are not evenly distributed across macro-areas. In the table in (57), the number of patterns is given for each macro-area and type combination. The Muna-type is mainly found in Papuanesia ( $\mathrm{n}=19$ ) and only rarely attested in Eurasia ( $\mathrm{n}=3$ ). Similarly, the Hebrew-type is mostly restricted to Eurasia ( $\mathrm{n}=9$ ) except for one outlier in Papuanesia. The patterns in Africa and North America show no such interaction.

| Interaction between infix interaction type and area in DoCoMD |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: |
| Macro-Area | Muna-type | Hebrew-type | No Interaction | Total |
| Papuanesia | 19 | 3 | 14 | 36 |
| Eurasia | 1 | 9 | 3 | 13 |
| North America | 0 | 0 | 2 | 2 |
| Africa | 0 | 0 | 4 | 4 |
| Total | 20 | 12 | 23 | 55 |

The sample here also shows a bibliographical bias. The interaction between infixation and root domain constraints has been described in much more detail for Austronesian languages, which make up the majority of the languages with infixation in DoCoMD $(\mathrm{n}=20)$, as shown in (58). Afro-Asiatic is the second largest top-level family here ( $\mathrm{n}=5$ ), whereas all other included top-level families show no more than one entry per family.

(58) | Genealogical distribution of infixation in DoCoMD |  |
| :--- | :--- |
| Top-level family | \# of languages |
|  | 20 |
| Afro-Asiatic | 5 |
| Austroasiatic | 1 |
| Misumalpan | 1 |
| Nakh-Dagestanian | 1 |
| Nuclear Torricelli | 1 |
| Nuclear Trans New Guinea | 1 |
| Sino-Tibetan | 1 |
| Siouan | 1 |
| Total | 32 |

The table in (59), shows an interesting difference between the Muna-type of interaction and the Hebrew-type of interaction. Whereas the Muna-type is largely restricted to the Austronesian language family ( $\mathrm{n}=19$ ) with only one outlier in Afro-Asiatic, the Hebrew type is spread out across different language families of Eurasia and Papuanesia. It occurs in Austronesian, Afro-Asiatic, Nakh-Dagestanian, Sino-Tibetan, Austroasiatic and Nuclear Toricelli. This is relevant because the Hebrew type has a lower incidence in general, but its genealogical spread makes it clear that this is not an isolated case.
(59) Interaction between infix interaction type and top-level family in DoCoMD

| top-level family | Muna-type | Hebrew-type | None | Total |
| :--- | ---: | ---: | ---: | ---: |
| Austronesian | 19 | 1 | 5 | 26 |
| Afro-Asiatic | 1 | 4 | 5 | 10 |
| Nakh-Dagestanian | 0 | 1 | 0 | 1 |
| Siouan | 0 | 0 | 1 | 1 |
| Nuclear Trans New Guinea | 0 | 0 | 9 | 0 |
| Sino-Tibetan | 0 | 2 | 0 | 2 |
| Austroasiatic | 0 | 2 | 2 | 4 |
| Misumalpan | 0 | 0 | 1 | 1 |
| Nuclear Torricelli | 0 | 2 | 0 | 2 |
| Total | 20 | 12 | 23 | 55 |

I have thus shown that both the Muna-type of interaction between root domain constraints and the Hebrew-type are robustly attested, whereas the Muna'-type and the Hebrew' do not show up in the crosslinguistic sample. In the following two sections, I show that both the Austronesian and the Afro-Asiatic language families do not show a completely uniform behaviour when it comes to interaction type. In section 3.7.3, I will provide additional data from other language families.

### 3.7.1 Austronesian Languages

Zuraw \& Lu (2009) collect several similar patters across Austronesian languages. They state that in all of these cases a repair strategy, avoiding two labial consonants, applies to the infix+root complex and the discontiguous root. Therefore these infixes can be grouped as infixes that undergo a root domain restriction, showing a Muna-like pattern. They also report a diverse variety of repair mechanisms and Tagalog as a language where the root domain constraint does not affect the infix.

Zuraw \& Lu (2009) only mention alternations of root segments with respect to infixation in two languages, Limos Kalinga (Austronesian, Philippines) and Northern Acehnese (Austronesian, Indonesia). These are cases of root dissimilation in order to satisfy root domain restrictions. It is clear from the result of the dissimilation (velar consonants or a coronal fricative) that these forms do not violate the root domain constraint against labial consonants. Therefore the root domain constraint still applies to the root domain in infixed forms. An example from Acehnese, where labial consonants in roots become a coronal fricative if preceding the infix $\langle\mathrm{um}\rangle$, is given in (60). ${ }^{25}$ This could be analyzed using Hierarchical Morphoprosodic Structure and both relativized versions of the root

[^41]domain constraint $\operatorname{OCP}(\mathrm{Lab})_{\sqrt{\omega}, \mathrm{D}}$ and $\mathrm{OCP}(\mathrm{Lab})_{\sqrt{\omega}, \mathrm{DD}}$ ranked above an IDENT constraint．
（60）Infix－triggered root dissimilation in Northern Acehnese（Durie 1985：33）
a．［s＊unãfoh］
／p〈um〉ałoh／
〈INTR $\rangle$ eat
＇to eat＇
b．［s＊umũpreh］
／p〈um〉upreh／
$\langle$ INTR $\rangle$ wait
＇to wait＇

All in all，Austronesian languages show mostly a Muna pattern．All twenty Austronesian infixing languages are given in（61）．The only pattern that is exceptional is Tagalog，which shows an interaction of the Hebrew－type．${ }^{26}$

[^42](61) Austronesian root domain constraints and infixation

| Language | Type |
| :--- | :--- |
| Aborlan Tagbanwa | Muna |
| Kanakanavu | Muna |
| Kapampangan | Muna |
| Kavala | Muna |
| Kimaragang | Muna |
| Limos Kalinga | Muna |
| Lolak | Muna |
| Mayrinax Atayal | Muna |
| Muna | Muna |
| Northern Acehnese | Muna |
| Paiwan | Muna |
| Palauan | Muna |
| Puyuma | Muna |
| Sarangani Blaan | Muna |
| Sarangani Manobo | Muna |
| Sediq | Muna |
| Tagalog | Hebrew |
| Thao | Muna |
| Timugon Murut | Muna |
| Toba Batak | Muna |
|  |  |

In the rest of this section, I will describe the data from Tagalog (Austronesian, Philippines) in order to contrast it with the patterns already described above and to show that inside Austronesian, we find Hebrew-like patterns as well as Muna-like patterns.

Tagalog shows a Hebrew-like pattern for some root-domain restrictions. In Tagalog, labial consonants only very rarely cooccur in native non-reduplicated roots. Nevertheless, if the actor-focus infix $\langle u m\rangle$ is infixed after labial consonants, the resulting forms do not usually show effects of a repair operation, as shown in (62). This is crucially different from the patterns reported in other Austronesian languages. The present approach would thus posit a different ranking for Tagalog, where only the version of the root domain constraint relativized to the narrow domain of the prosodic root word, ${ }^{*} \mathrm{BM}_{\sqrt{\omega}, \mathrm{DD}}$, outranks faithfulness constraints.
（62）Tagalog um－infixation without repairs（Zuraw \＆Lu 2009：199）
a．$\quad \mathrm{p}\langle u m\rangle$ ili
〈ACT．FOC〉－choose
＇choose＇
b．$b\langle u m\rangle$ ukas
〈ACT．FOC〉－open
＇open＇

Note，however，that this holds true only for labial obstruents，／p／and／b／．If the initial consonant is a labial sonorant $/ \mathrm{w} /$ or $/ \mathrm{m} /$ ，a paradigm gap results．This systematic gap has been experimentally confirmed in a nonce－word task by Orgun \＆Sprouse（1999）． Orgun \＆Sprouse（1999）tested English loan words into Tagalog for acceptance by native speakers．All words were accepted with the infix $\langle u m\rangle$ ，except loans that start in $/ \mathrm{m} /$ or $/ \mathrm{w} /$ ．Differing behavior for different kinds of root domain constraint is expected under the present approach，since different relativized markedness constraints can be ranked above or below faithfulness constraints．${ }^{27}$
（63）Tagalog um－infixation results in paradigm gaps（Orgun \＆Sprouse 1999：206）
a．${ }^{*} \mathrm{~m}\langle u \mathrm{um}\rangle$ eri
$\langle$ ACT．FOC $\rangle$－marry
＇marry＇
b．${ }^{*}$ w $\langle u m\rangle$ ejl
〈ACT．FOC〉－wail
＇wail＇

## 3．7．2 Afro－Asiatic Languages

A similar pattern in Arabic is explicitly linked to the Domain Problem in McCarthy（1979）． The reflexive infix $\langle t\rangle$ is ignored by the OCP in Arabic．It neither undergoes it nor does it block its application to the root．The infix thus does not enter the domain of the OCP（C） and the root stays a domain if infixed．Arabic shows interaction of the Hebrew－type．

Arabic roots cannot have two identical consonants as two non－final adjacent members，just as Hebrew roots．This constraint does not apply in derived environments．The first two consonants of a word can be identical if at least one of them is a prefix．

[^43](64) $\quad \mathrm{OCP}(\mathrm{C})$ in Arabic roots (McCarthy 1979:263,264,265)
a. katab
write.PST
'he wrote'
b. ${ }^{* *}$ sasam
c. ${ }^{* *}$ dadraj
d. ${ }^{* *}$ tatak
(65) No OCP (C) in Arabic words (Aya Al-Ghanem, p.c.)
a. ta-ta-kallam (cf. **tatak) 2SG-REFL-talk 'you converse'
b. ta-truku (cf. **tatak) 2SG-leave 'you leave'

The infix $\langle\mathrm{t}\rangle$ in Arabic is cognate to other Semitic infixes like the Hebrew one mentioned above. It is accompanied by a prefixed exponent /Ri/-. Similarly, the meaning of these reflexive forms is often not fully transparent.
(66) Reflexive infix in Arabic (McCarthy 1979:209)
a. kataba
write.PST
'he wrote'
b. Pi-k $\langle\mathrm{t}\rangle$ ataba
$\mathrm{REFL}_{i}-\left\langle\mathrm{REFL}_{i}\right\rangle$ write
'he copied'

Crucially, the $\langle\mathrm{t}\rangle$ infix does not interact with the $\mathrm{OCP}(\mathrm{C})$ on roots. If a root starts with a /t/, the reflexive form on the surface seems to violate the $\mathrm{OCP}(\mathrm{C})$. There is no repair strategy applied here to resolve this issue.
(67) Infix is not subject to $\operatorname{OCP}(\mathrm{C})$ (Aya Al-Ghanem, p.c.)
a. Pi-t $\langle\mathrm{t}\rangle$ abaSa
$\operatorname{REFL}_{i}-\left\langle\mathrm{REFL}_{i}\right\rangle$ comply
'he complied with'

Additionally, the $\mathrm{OCP}(\mathrm{C})$ still applies in root domains in infixed forms. There are no roots that change their second consonant in infixed form such that it would be identical to
their first consonant. This is shown in (68).
(68) Non-existing infixed forms in Arabic
a. ${ }^{* *}$ ?i-k $\langle t\rangle$ akab
b. ${ }^{* *}$ Pi-k $\langle\mathrm{t}\rangle$ akal

Arabic is thus another instance of the Hebrew pattern. The $\mathrm{OCP}(\mathrm{C})$ does not apply to the infix and no repair strategy is applied to get rid of this violation. The root stays a domain for $\mathrm{OCP}(\mathrm{C})$ after infixation and no alternations are found that would indicate that the $\operatorname{OCP}(\mathrm{C})$ is relaxed in this context. An analysis using Hierarchical Morphoprosodic Structure would thus be analogous to the analysis of Hebrew with only a root domain constraint relativized to the narrow domain of the prosodic root word $\operatorname{OCP}(\mathrm{C})_{\sqrt{\omega}, \mathrm{DD}}$ ranked above the corresponding faithfulness constraint.

Another Semitic language with a reflexive $\langle\mathrm{t}\rangle$ infix is the extinct language Akkadian. Akkadian is different from Hebrew and Arabic in that the infix undergoes various repair strategies in order to satisfy an OCP constraint (cf. Lubowicz 2010). In Akkadian, a constraint virtually identical to the ones in Arabic and Hebrew is operative in the root domain (McCarthy 1986). More specifically, no two coronal consonants can occur in a root as adjacent radicals (cf. Reiner 1965:49). This is relevant, since the infix also consists of a coronal consonant. ${ }^{28}$ I will refer to this constraint as OCP(Cor). This is different from the $\operatorname{OCP}(\mathrm{C})$ used before. The $\mathrm{OCP}(\mathrm{C})$ penalizes fully identical consonants, whereas the $\mathrm{OCP}(\mathrm{Cor})$ bans any consonants that have a coronal place of articulation independent of other feature specifications.

Interestingly, even though the $\langle\mathrm{t}\rangle$ infix does not block the OCP(Cor) constraint, it undergoes the $\operatorname{OCP}($ Cor $)$ itself in that it alternates in form and locus. (McCarthy 1979; Lubowicz 2010). ${ }^{29}$ In Akkadian, the reflexive stem is regularly formed by infixing - $\langle$ ta $\rangle$ after the first (C)V-sequence of the root. The infix often shows up as $\langle\mathrm{t}\rangle$ due to phonological processes. ${ }^{30}$

[^44]（69）Akkadian reflexive infix（Lubowicz 2010：14）
a．i－ $\int\langle$ ta $\rangle$ kan
3SG－〈REFL〉settle
＇he has settled＇（cf．Jaka：n－um＇to settle＇）
b．mi $\langle\mathrm{t}\rangle$ gur－um
〈REFL〉agree－INF
＇to agree with one another＇（cf．magarr－um＇to agree＇）
c．a〈ta〉ppul－um
〈REFL〉anwer－INF
＇to answer，pay repeatedly＇（cf．apa：l－um＇to answer＇）

However，if the root begins with a coronal consonant and the first two consonants in the root domain are separated by a vowel，／ta／occurs as a prefix．Lubowicz（2010）interprets this as a repair strategy to avoid an $\mathrm{OCP}($ Cor $)$ violation in the root domain．The $\mathrm{OCP}(\mathrm{Cor})$ cannot be violated by an infix inside the root．In a Hierarchical Morphoprosodic Structure analysis，both markedness constraint relativized to the prosodic root word，OCP（Cor）$\sqrt{\sqrt{\omega}}$ ，DD and $\operatorname{OCP}(\mathrm{Cor})_{\sqrt{\omega}, \mathrm{D}}$ ，would need to outrank a Linearity constraint in order to allow for this repair mechanism．
（70）Akkadian prefixed／t／（Lubowicz 2010：15）
a．t－its＇but－um，＊ts＇i $\langle\mathrm{t}\rangle$ but－um
REFL－touch－INF
＇to touch one another＇（cf．ts＇aba：t－um＇to quarrel＇）
b．t－isk＇ar－i：，$\quad * t i\langle t\rangle$ k＇ar－i：
REFL－speak－IMP．F
＇pronounce forever！＇（cf．sak＇a：r－um＇to speak＇）
c．t－izkur－um＊zi $\langle\mathrm{t}\rangle$ kur－um
REFL－speak－INF
＇to speak＇（cf．zaka：r－um＇to speak＇）

Lubowicz（2010）also describes a second repair strategy．In some forms the infixed／t／ would be adjacent to the root－initial coronal．This is due to syncope in most cases．In this context，an assimilation process applies．This avoids the OCP（Cor）violation，since geminates count as one consonant for this purpose．
(71) Akkadian repair gemination (Lubowicz 2010:15)

```
a. i-d\langle:a\ranglemik' *i-d\langleta\ranglemik'
    3SG-REFL-improve
    'he has improved' (cf. dama:k'-um 'to improve')
b. i-s\langle:a\rangle-xap *i-s \ta\ranglexap
    3SG-REFL-cover
    'he has covered' (cf. dama:k'-um 'to cover')
c. i-t'<:>a-rad *it'\langleta\ranglerap
    3SG-REFL-send
    'he has sent' (cf. t'ara:d-um 'to send')
```

All in all, Akkadian shows a pattern very different from Arabic and Hebrew. The OCP(Cor) applies to the root but it also applies to the infix, i.e. the infix undergoes a repair mechanism to satisfy the OCP(Cor) constraint. Akkadian thus shows a Muna-like pattern. The AfroAsiatic languages therefore show different patterns. Akkadian is different from Hebrew and Arabic varieties. Additionally, Mupun (Chadic; Nigeria) shows no interaction whatsoever. The verbal plural infix $\langle a\rangle$ cannot interact with any constraint on monomorphemic domains, since no root domain constraints apply to low vowels in Mupun.
(72) Infixation Interaction types in Afro-Asiatic languages in DoCoMD

| Language | Type |
| :--- | :--- |
| Akkadian | Muna |
| Hebrew | Hebrew |
| Muscat Arabic | Hebrew |
| Mupun | None |
| Standard Arabic | Hebrew |

### 3.7.3 Other Language Families

A different example comes from Hunzib (Nakh-Dagestanian; Georgia, Russia). Hunzib shows a Hebrew patttern. Even though roots can be maximally bisyllabic (van den Berg 1995:27; Kalin 2021:3), infixes can create trisyllabic words. Consider the data in (73). Native roots in Hunzib can be maximally disyllabic, as in the roots shown in (73-a) and (73-b). Trisyllabic roots do not exist, cf. (73-c), where the form is possible only if morphologically complex.
(73) Attested and impossible forms in Hunzib (van den Berg 1995:28-31; Kalin 2021:3)
a. Ríyu
'mother'
b. k'išáa
'play'
c. Pis-ná-la-s
'siblings (genitive)' (cf. **?isnálas)

This becomes crucial when looking at the infix $\langle a\rangle$. It expresses pluractionality and is placed before the final consonant. If it occurs after a vowel that would receive stress in the uninfixed form, hiatus is resolved by glide insertion, i.e. the infix adds a whole syllable inside the root.
(74) Infixes create trisyllabic forms (van den Berg 1995:334; Kalin 2021:4)
a. šóše
'bandage'
b. šo〈wá〉še
'bandage (plural)'

This means that infixed forms violate the root domain constraint. Crucially, there are no roots that vary if infixed and would add a root syllable if infixed. Therefore the infix neither joins the root domain nor does it destroy the domain of the simple root. We can thus classify these data as a Hebrew pattern. An analysis in the approach presented here would rank a markedness constraint allowing maximally bisyllabic roots above a faithfulness constraint against deletion: ${ }^{*} \sigma \sigma \sigma_{\sqrt{\omega}, \mathrm{DD}} \gg$ MAx.

One intriguing difference is that this pattern relies on a root domain constraint that references syllables, i.e. derived prosodic structure. This is an argument against a MSCbased account, since syllables are not present in the underlying structures under most approaches to MSC, since they constitute predictable information (cf. also chapter 2).

It needs to be said that infixation and root domain constraints do not always interact. In some cases, infixation never violates a root domain constraint. In Semelai (Austroasiatic, Malaysia) for example, the coda of the penultimate syllable of a monomorphemic root can only be from a very restricted set of consonants, namely $/ \mathrm{m} / \mathrm{h} / \mathrm{n} / \mathrm{l} / \mathrm{y} / \mathrm{l} / \mathrm{l} /$, or $/ \mathrm{r} /$ (Kruspe 2004:53). One infix that appears in that position is the causative marker $\langle r\rangle$. Since the infix can never violate the root domain constraint, these data do not tell us anything about possible interactions between infixation and root domain constraints. However, Semelai also features a form of infixing reduplication, termed coda copy by Kruspe (2004), which
can violate the root domain constraint. The coda consonant of the final root syllable is copied and inserted as the coda consonant of the penultimate syllable. Final syllables allow a greater variety of consonants than penultimate syllables. Therefore, faithful copying will create penultimate codas that are not available for simple roots. For further discussion of the Semelai pattern see chapter 2. Semelai thus shows both no interaction and a Hebrew-type interaction, depending on the infixing process. This is expected under Richness of the Base, where both infix forms are allowed underlyingly, those that do violate root domain constraints and those that do not.

A third example of the Hebrew type comes from Yeri (Nuclear Toricelli; Papua New Guinea). Here, superheavy syllables and non-initial closed syllables are not allowed in monomorphemic words but can occur if either a diphthongal infix or a coda consonant infix occurs (Wilson 2017:70). Infixation in Yeri can thus create structures that are not attested in simple roots. An overview about infixation interaction outside of Austronesian and Afro-Asiatic is given in (75).
(75) Infixation interaction in other language families

| Language | Top-level family | Type |
| :--- | :--- | :--- |
| Hunzib | Nakh-Dagestanian | Hebrew |
| Lakhota | Siouan | No Interaction |
| Dani | Nuclear Trans New Guinea | None |
| Pinging Mandarin | Sino-Tibetan | Hebrew |
| Semelai | Austroasiatic | None, Hebrew |
| Ulwa | Misumalpan | None |
| Yeri | Nuclear Torricelli | Hebrew |

### 3.8 Conclusion

In this chapter, I have shown that the empirical picture of interactions between root domain constraints and infixation has a systematic gap with serious theoretical implications. Infixation never destroys the root domain. Root domain constraints always apply to discontiguous roots created by infixation. In other words, a root domain constraint never ceases to apply to the simple root in infixed forms. This contrasts with the variability by which infixes can either become part of the domain of a root domain constraint (Muna pattern) or not (Hebrew pattern). This is unexpected from a procedural point of view, since infixation should be able to bleed the repair strategy for the root domain constraint. I have shown that less restrictive procedural approaches like SPE and Stratal OT therefore face a severe overgeneration problem, whereas more restrictive approaches like Harmonic Serialism and Output-Output Faithfulness cannot account for the full empirical
picture. Instead I have proposed an account based on Hierarchical Morphoprosodic Structure. Its main advantage is that it can be restricted to exclude the unattested Muna' pattern by introducing a simple fixed ranking that is independently motivated by locality considerations. In effect, infixes are easy to differentiate in representations but not in procedures. This is because infixes are representationally special in that they linearly interrupt a monomorphemic domain.

## Chapter 4

## Trigger Asymmetries in Vowel Harmony

Clements (1984, 1985) was the first to note that a rule-based analysis of a certain kind of vowel harmony can be substantially simplified if constraints on underlying representations are allowed. In this kind of trigger asymmetry, a certain vowel never occurs as the trigger of harmony and never shows up in morphologically underived environments but nevertheless is licit if derived by vowel harmony. This can be easily modeled if a constraint on underlying representations bans the asymmetric non-trigger from the input and a vowel harmony rule derives the same vowel in the correct derived contexts. Clements also correctly pointed out that such a pattern poses a challenge for an approach based on surface-oriented constraints: if a constraint-based approach does not distinguish between triggers and targets of vowel harmony, the asymmetric non-trigger should be able to surface as either in vowel harmony contexts. A concise definition is provided in (1).
(1) Asymmetric non-trigger

An asymmetric non-trigger is a type of vowel that occurs as a result of vowel harmony but neither as a trigger of vowel harmony nor in underived environments.

I will illustrate the trigger asymmetry problem ${ }^{1}$ more concretely with the Akan data from the Asante Twi dialect discussed in Clements (1984, 1985). In Akan (Atlantic-Congo; Ghana, Togo), a low vowel /a/ is raised to $[\partial]^{2}$ if it directly precedes a high vowel. In (2-a), /a/ cannot occur, since the form ends in /i/, only [kəri] is attested. Note that raising

[^45]applies both inside the root and across a prefix-root boundary. As shown in (2-b), a prefix cannot surface with the vowel [a] before a root with /i/ as its leftmost vowel, i.e. a form like *[adi] is unattested. In all other cases it surfaces faithfully as $/ a /$, as shown in the examples (2-c,d).
(2) Distribution of [a] and [ə] in Akan (Clements 1984:325,326)
a. kəri, **kari
'to weigh'
b. w-ə-di, *w-a-di
'he has eaten'
c. bisa, ${ }^{* *}$ bisə
'to ask'
d. kasa, ${ }^{* *}$ kəsə
'to speak'

Raising of /a/ is part of a more general pattern of vowel harmony, where [-ATR] vowel become [+ATR] before other [+ATR] vowels. (cf. Clements 1984:326,fn 2). Such patterns are given in (4). As can be seen, mid and high vowel assimilate to a root vowel in the feature $[ \pm$ ATR $]$. This happens to both prefixes and suffixes. The prefix in $(3-\mathrm{a}, \mathrm{b})$ shows up as [e] before the $[+\mathrm{ATR}]$ root $/ \mathrm{bu} /$ in ( $3-\mathrm{a}$ ) but as $[\varepsilon]$ before the $[-A T R]$ root $/ \mathrm{bv} /$ in (3-b). Similarly, the suffix in (3-c,d) shows up as [i] after the [+ATR] root /tu/but as $[\mathrm{I}]$ following the $[-\mathrm{ATR}]$ roots [tv] and [kasa]. The pattern is thus root-controlled (but cf. Casali 2012).
(3) Akan regular harmony (Clements 1985:62,78)
a. e-bu-o
'nest'
b. $\varepsilon$-br-o
'stone'
c. o-be-tu-i
'he came and dug it'
d. $\quad \mathrm{o}$-be-tz-I
'he came and threw it'
e. o-kasa-i
'he spoke'

The vowel inventory is given in (4). As shown in (3), high, mid and low vowels can trigger vowel harmony. These are underlined in (4). The only vowel that does not occur as a
trigger is [ə]. This vowel only occurs if derived by vowel harmony.
(4) Akan vowel inventory (Clements 1985:57)

|  |  | Front | Back |
| :---: | :---: | :---: | :---: |
| high | +ATR | $\underline{1}$ | $\underline{\underline{u}}$ |
|  |  | $\downarrow$ | , |
|  | -ATR | I | $\underline{\square}$ |
| mid | +ATR | $\underline{\text { e }}$ | - |
|  |  | 1 | $\downarrow$ |
|  | -ATR | $\underline{\varepsilon}$ | $\bigcirc$ |
| low | +ATR |  | ә |
|  | -ATR |  | $\uparrow$ |

Since the pattern is root controlled, any trigger of vowel harmony has to be situated in the root. Crucially, a root with [ $[\exists$ ] as its only vowel is impossible. Similarly, [ $\rho$ ] cannot occur in a root without a following high vowel. ${ }^{3} / \partial /$ thus cannot trigger $[+$ ATR $]$ harmony, even though it is a [+ATR] vowel. These data are easily explained in an approach that employs constraints on the underlying representation of segments. / / can be easily excluded from triggering harmony if it is banned from underlying representations and only surfaces as the result of vowel harmony.
(5) Derivation table for asymmetric triggers in Akan vowel harmony

| Input: | ว-kəSə-I | --kasa-I | w-ə-di | w-a-di |
| :--- | :--- | :--- | :--- | :--- |
| $*_{\partial}$ | - | n.a. | - | n.a. |
| Vowel Harmony | - | okasai | - | wədi |
| Output | - | okasai | - | wədi |

The same data pose a problem for a simple parallel OT analysis. If / / can surface in order to satisfy a cover constraint for vowel harmony, the ranking of the two must be HARMONY $\gg$ *. Under this ranking, however, we would expect $[ə]$ to occur whenever other candidates violate the vowel harmony constraint, even if [ $\partial$ ] occurs in a triggering position. Simple parallel OT does not distinguish between the trigger and the target of vowel harmony without additional stipulations. In the tableau in (6), candidate ( $6-\mathrm{b}$ ) with a triggering /ə/ in the root would become optimal, even though it is not the attested form. ${ }^{4}$

[^46]The problem with candidate ( $6-\mathrm{c}$ ) is that the root vowel assimilated to the affix vowel in [ $\pm \mathrm{ATR}$ ], thereby violating the vowel harmony constraint that requires root control, i.e. affixes assimilating to roots.
(6) Failure of simplistic SPOT analysis

| $\mathrm{I}: / \partial /-\mathrm{k} \partial \mathrm{s} \partial /-/ \mathrm{I} /$ | HARMONY(ATR) | $* \partial$ | $\operatorname{IDENT}(\mathrm{ATR})$ |
| :---: | :---: | :---: | :---: |
| a. /ə/-kəsə/-/ı/ | $*!$ |  |  |
| b. /o/-kəsə/-/i/ |  | $*$ | $*$ |
| Bc. /ว/-kasa/-/ı/ | $*!$ |  | $*$ |

In this chapter, I will argue that parallel OT with RotB can explain these data if provided with the correct representation. I will argue that a representational approach based on the independently proposed Headed Spans Theory (McCarthy 2004) is also able to derive the data by directly restricting certain vowels from triggering harmony. Headed Span Theory allows us to representationally distinguish between trigger and target as a head and a non-head of the span. Constraints on underlying representations are not needed. I will also make the stronger claim that these representations allow us to maintain a surface-oriented view of Optimality Theory, formalized as Output Drivenness (Tesar 2013). This will prove to be a advantage of the present analysis over previous approaches.

The remainder of this chapter is organized as followed. In section 4.1, I will introduce Headed Spans Theory. This will be followed by two representative case studies of asymmetric non-triggers in vowel harmony from Assamese and Päri in section 4.2 and 4.3, respectively. While the analysis of Assamese will introduce the more general part of the Headed Spans analysis, the analysis of Päri will show that this analysis can also be expanded to include morphologically conditioned exceptions to the asymmetric non-trigger status of some vowels. Section 4.4 introduces the notion of Output Drivenness and discusses previous approaches to asymmetric non-triggers. I will show that none of the alternative approaches can maintain Output Drivenness in contrast to an analysis based on Headed Spans. After discussing crosslinguistic data in section 4.5, I will conclude in section 4.6.

### 4.1 Headed Spans Theory

Headed Spans Theory (McCarthy 2004) is part of a larger class of approaches that try to adjust autosegmental representation for use with Optimality Theory. The main difference here is that autosegmental association to a feature is assumed to be asymmetric. One segment serves as the designated head of the so-called autosegmental span, whereas all other associated segments are dependents or non-heads. Other theories of this family include Optimal Domain Theory (Cole \& Kisseberth 1994; Cassimjee \& Kisseberth 1998) and

Headed Feature Domains (Smolensky et al. 2006). ${ }^{5}$ I will mark the head by underlining the segment as well as associating it by a bold association line in the autosegmental structure, as shown in (7). Note that if a feature is only connected to one segment, this segment must serve as the head of the span, cf. (7-a). Additionally, I follow McCarthy (2004) in enclosing the feature span, i.e. all segments associated with a certain relevant feature, in brackets.
(7) Simple and complex headed spans

(i)
( $\underline{\varepsilon}$
a.

+ ATR

$$
\left(\begin{array}{ll}
\mathrm{i} & \mathrm{e}
\end{array}\right)
$$

b.

The main innovation of a headed spans approach is the notion of a head of a feature span. Since the head is present in the representation, constraints can refer to the features it is associated to. These markedness constraints can be of several types. They can ban or require a certain feature value for a head of a span with a certain feature value. They can also require a segment with a certain feature value to head a span with a certain feature value. For the following discussion, the former class of constraints will be crucial. ${ }^{6}$ In (8), a concrete example is given, which will be used in the case studies later. The constraint bans mid vowels from heading [+ATR] spans.
(8) Featural markedness constraints on span heads
a. $\quad$ * $\operatorname{SpanHEAD}(\mathrm{e}, \mathrm{o})(+\operatorname{ATR})(=* \operatorname{SpHd}(\mathrm{e}, \mathrm{o}))$

Count one violation for every [+ATR] span that is headed by a short mid vowel.

Further constraints are needed for a complete analysis of a vowel harmony pattern. Headed Spans Theory includes markedness constraints to drive harmony and to derive directionality as well as a faithfulness constraint to penalize feature changes. The generalized markedness scheme that acts as a harmony driver is a constraints against adjacent spans,

[^47]*AdjacentSpan(F). Note that this is not a new constraint, since it effectively penalizes adjacent features on a tier, similar to the ObligatoryContourPrinciple(F) constraint proposed by Leben (1973) and adapted into OT by Myers (1997). The formulation of a concrete example of this constraint type is given in (9-a). Directionality is derived by a constraint that requires a span to be headed by its leftmost or rightmost segment. ${ }^{7}$ Note that the directionality constraints are analogous to the constraints on span heads in (8), the only difference being the restrictions on position or on features, respectively. These constraints can in principle also restrict the head to bear other properties, e.g. being stressed, as I will assume in the analysis of Päri in section 4.3.
(9) Further constraints in headed spans theory
a. *AdJacentSpans(ATR) (=AS)

Count one violation for each pair of adjacent [ $\pm$ ATR] spans
b. SpanHeadRight(+ATR) (SH-R)

Count one violation for every [ + ATR] span that is not headed by its leftmost segment.
c. FaithfulnessHeadSpan(+ATR) (=FHS)

Count one violation for any corresponding pair of an input segment and an output segment, where the input segment is the head of a [+ATR] span and the output segment is not.

Headed Spans Theory was originally motivated by its ability to solve empirical problems of competing approaches to vowel harmony. Standard theories of assimilation in Optimality Theory, most prominently the Agree(F) constraint, run into the so-called sour grapes problem, when a blocking segment occurs. Since such a constraint only requires a pair of locally adjacent segments to agree in a feature, the constraint does not differentiate two candidates with the same number of non-agreeing pairs in different positions. Faithfulness constraints then favor the more faithful candidate. Assimilation becomes a matter of all-or-nothing. It only occurs if it can affect all targets without any intervening blockers, otherwise it does not apply at all.

A concrete example is given by McCarthy (2004), the form /pəŋ/-/awasan/ [pəŋãw̃ãsan] 'supervision' from Johore Malay (Austronesian, Malaysia), where the feature nasal assimilates from left to right and /s/ blocks nasal harmony. An account based on Agree(F)

[^48]constraints cannot distinguish between the maximally faithful output [pəŋawasan] and the attested output [pəŋãw̃ãsan]. Both incur the same number of violations, since both have three pairs of non-agreeing segments.

This problem does not occur in Headed Span Theory. The solution is a constraint that is satisfied only if the blocker /s/ heads a [-nasal] span. Since spans heads of [ $\pm$ nasal] spans have to be initial in their span, this constraint is necessarily violated by [(pə)(므)(awasa)(n)] but not by $[(\underline{p} ə)(\underline{\underline{n}} \tilde{a} \tilde{w} \tilde{a})(\underline{\underline{s}} \mathbf{a})(\underline{n})]$. Therefore, Headed Span Theory solves the sour grapes problem. ${ }^{8}$

Headed Spans Theory was thus independently proposed to solve pathologies of existing approaches to assimilation in Optimality Theory. This is accomplished by its asymmetric representation of autosegmental spans, where one segment is the designated head. Constraints that refer to this head status solve the sour grapes problem. In the following two representative case studies, I will show that the same assumptions allow for an analysis of the asymmetric non-trigger problem.

### 4.2 Only High Triggers in Assamese

Vowels in Assamese (Indo-European, India) harmonize for the feature [+ATR] (Mahanta 2008). The harmony is dominant-recessive and operates from right to left. The vowel inventory of Assamese is given in (10). Arrows indicate direction of change under vowel harmony. Triggers are underlined. As shown, only high vowels trigger [+ATR] harmony. Low vowels do not participate in regular harmony. [+ATR] mid vowels $[\mathrm{e}, \mathrm{o}]$ serve as asymmetric non-triggers. The [-ATR] counterpart of /i/ is simply missing from the vowel inventory.

Assamese vowel inventory (Mahanta 2008:58)

|  |  | Front | Back |
| :---: | :---: | :---: | :---: |
| high | +ATR | $\underline{i}$ | $\underline{\mathrm{u}}$ |
|  |  |  | $\uparrow$ |
| mid | -ATR |  | $v$ |
|  | +ATR | e | $o$ |
|  |  | $\uparrow$ | $\uparrow$ |
| low | -ATR | $\varepsilon$ | $\supset$ |
|  |  |  | $a$ |

As shown in (11), vowel harmony is regressive. If a [-ATR] vowel occurs before a [+ATR] vowel, it changes into its $[+\mathrm{ATR}]$ counterpart. High vowels (cf. (11-a)) as well as mid

[^49]vowels (cf. (11-b)) can undergo harmony. The root vowel /v/ becomes [u] when the root $/ \mathrm{gvl} /$ precedes the suffix $/ \mathrm{i} /$, resulting in the surface form [guli]. In the reverse situation, i.e. if a [-ATR] vowel follows a [+ATR] vowel, the harmony does not apply. This is illustrated for [+ATR] high vowels before [-ATR] mid vowels in (11-d) and before [-ATR] mid vowels in (11-c). The forms $/ \mathrm{p}^{\mathrm{h}} \mathrm{ur} /-/ v /$ and $/ \mathrm{p} v \mathrm{r} /-/ \varepsilon /$ both surface faithfully. This implies that both directionality and dominance are needed to describe the vowel harmony pattern. If it were dominant-recessive bidirectional [+ATR] harmony, we would expect progressive vowel harmony in (11-c) and (11-d) yielding $*\left[p^{\mathrm{h}}\right.$ uru] and $*\left[b^{\mathrm{h}}\right.$ ute], respectively. If the harmony system was purely regressive, we would expect regressive assimilation in these cases, resulting in $*\left[\mathrm{p}^{\mathrm{h}} \psi \mathrm{rv}\right]$ and $*\left[\mathrm{~b}^{\mathrm{h}} v t \varepsilon\right]$.

Right-to-Left [+ATR] harmony in Assamese (Mahanta 2008:7,91,94, Mahanta 2012:1112)
a. /gvi/-/i/ $\rightarrow$ [guli] 'to mix'
b. /pet/-/u/ $\rightarrow$ [petu] 'pot bellied'
c. $/ \mathrm{p}^{\mathrm{h}} \mathrm{ur} /-/ v / \rightarrow$ [ $\mathrm{p}^{\mathrm{h}}$ urv] 'travel, roam (1.PRS)'
d. $/ \mathrm{b}^{\mathrm{h}} \mathrm{ut} /-/ \varepsilon / \rightarrow\left[\mathrm{b}^{\mathrm{h}} \mathrm{ut} \mathrm{\varepsilon}\right] \quad$ 'ghost (ERG)'
e. $/ \mathrm{pvr} /-/ \varepsilon / \rightarrow[\mathrm{p} v \mathrm{r} \mathrm{\varepsilon}] \quad$ 'fall (3.PRS)'
f. $/ \mathrm{k} \partial \mathrm{r} /-/ \mathrm{\rho} / \rightarrow[\mathrm{k} \supset \mathrm{r}] \quad$ 'do (2.PRS)'
/e/ and /o/ however do not occur in isolated native roots. They never trigger vowel harmony at all. Mahanta (2008:77) therefore assumes that [+ATR] mid vowels are not present in the lexicon and only occur if derived by vowel harmony. As shown in (12-a), [ +ATR$]$ suffixes are banned because they would otherwise be expected to trigger a change from [-ATR] to $[+\mathrm{ATR}]$ in the preceding vowel of the root. Similarly, if a $[+\mathrm{ATR}]$ root would exist, we would expect it to surface faithfully in front of [-ATR] suffixes (12-b) and in isolation (12-c). Further examples of unattested roots are given in (13).
(12) Impossible roots and affixes

$$
\begin{array}{llll}
\text { a. } & \text { /gvl/-**/e/ } & \rightarrow & * *[\text { gule }] \\
\text { b. } & * * / \mathrm{p}^{\mathrm{h}} \text { or } /-/ v / & \rightarrow & * *\left[\mathrm{p}^{\mathrm{h}} \text { orv }\right] \\
\text { c. } & * * / \text { kor } / & \rightarrow & * *[\text { kor }]
\end{array}
$$

(13) Impossible isolated roots (Mahanta 2008:118)
a. ${ }^{* *}$ tero
b. ${ }^{* *}$ beton
c. ${ }^{* *}$ poxek
d. ${ }^{* *}$ olek $^{\text {h }}$

More abstractly, the problem for a surface-oriented approach can be described as follows. If [e] and [o] were prohibited from surfacing, we would not expect them to occur as the result of [+ATR] harmony. If [e] and [o] were allowed to surface in the contexts satisfying vowel harmony requirements, we would expect them to occur in isolated roots, where vowel harmony requirements are vacuously satisfied, and as triggers in affixes, where vowel harmony is satisfied by assimilation of the root to the affix.

The severity of the problem can be illustrated by a simplified OT analysis. ${ }^{9}$ Assuming that directionality, dominance, and the harmonizing feature are summarized in a cover constraint HARMONY(ATR), we could naively expect that ranking this constraint above a constraint against $[\mathrm{e}]$ an $[\mathrm{o}]$ would derive the pattern. However, the predicted pattern is different from the attested facts. The $[+\mathrm{ATR}]$ mid vowels would be generated as potential triggers. This is simply due to the fact that satisfying the Harmony constraint is more important than prohibiting [e] and [ o ]. If vowel harmony can be satisfied only by allowing [e] and [o], the [+ATR] mid vowel will surface, irregardless of its status as a trigger or target of vowel harmony. ${ }^{10}$
Failure of simplistic SPOT analysis

| I: | $/$ kor-o $/$ | HARMONY(ATR) | ${ }^{*} \mathrm{e}, \mathrm{o}$ | IDENT(ATR) |
| ---: | :--- | :---: | :---: | :---: |
| a. | kəro | $*!$ |  |  |
| -b. koro |  | $*$ | $*$ |  |
| Bc. | koro | $*!$ |  | $*$ |

I argue that the main problem with a simplistic OT analysis is the presumed representational symmetry of vowel harmony. Triggers and targets cannot be distinguished in the output representation. Headed Span Theory theory makes this possible by enriching autosegmental representation by the notion of a head. This allows us to easily state the generalization that [ + ATR] mid vowels cannot be the trigger of vowel harmony - or the head of a [+ATR] span. Crucially, the same constraint, given in (15), will also derive the absence of $[+$ ATR $]$ mid vowels in underived contexts.
*SpanHEAd (e,o)(+ATR):
Count one violation for every [+ATR] span that is headed by a [-high,-low] vowel.

Additional constraints are only needed to account for the more general properties of the

[^50]harmony. *AdJACENTSPANS(ATR) acts as a general harmony constraint by disallowing adjacent spans of the $[ \pm \mathrm{ATR}]$ feature. The directionality is ensured by the constraint SpanHeadRight(+ATR) that requires the head of [+ATR] span to be its rightmost segment. I will also follow McCarthy (2004) in assuming that dominance is triggered by a value-restricted faithfulness constraint, in this specific case FaithHeadSpan(+ATR).
(16) Further constraints on Assamese vowel harmony
a. *AdjacentSpans(ATR) ( $=$ *AS)

Count one violation for each pair of adjacent $[ \pm$ ATR $]$ spans.
b. SpanHead-Right(+ATR) (=SH-R)

Count one violation for an every [+ATR] span that is not headed by its final segment.
c. FaithfulnessHeadSpan (+ATR) $(=\mathrm{FHS}(+\mathrm{A}))$

Count one violation for a segment that is the head of a [+ATR]-Span in the input but not in the output.

Note that the markedness constraint on [+ATR] span heads has nothing to say about non-heads. For exactly this reason, derived $[\mathrm{e}]$ and $[\mathrm{o}]$ can occur in vowel harmony contexts. In the tableau in (17) the optimal candidate (17-b) satisfies the constraint since the $[+\mathrm{ATR}]$ mid vowel is not the head of any $[+\mathrm{ATR}]$ span. Therefore, harmony proceeds as it does in normal cases, spreading [+ATR] from right-to-left producing a non-head [+ATR] vowel. Candidate (17-a) is excluded, since vowel harmony has failed to apply, whereas candidate (17-c) cannot become optimal since harmony has applied in the wrong direction, violating both SpanHead-Right(+ATR) and FaithfulnessHeadSpan(+ATR). This evaluation shows that it is possible to rank *SpanHEAD $(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$ above the general harmony constraint without excluding $[\mathrm{e}]$ and $[\mathrm{o}]$ if derived by vowel harmony.

Derived [+ATR] mid vowels in Assamese as non-heads

|  | *SH(eo) | SH-R | FHS $(+\mathrm{A})$ | *AS | FHS |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | *! |  |
|  |  |  |  |  | * |
|  |  | *! | * |  | ** |

The high ranking of *SpAnHEAD $(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$ becomes crucial for keeping [+ATR] mid vowels in suffixes from triggering vowel harmony. If a $[+A T R]$ mid vowel in a suffix, such as -/e/ in (18), spreads its ATR feature to the left - as in candidate (18-c)—, this violates the *SpanHEAD $(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$ constraint because $[\mathrm{e}]$ now has become the head of a $[+\mathrm{ATR}]$ span. Due to the high ranking of this constraint, the candidate is excluded. This warrants an otherwise unexpected structure in candidate ( $18-\mathrm{b}$ ) to become optimal. The suffix vowel is neutralized to a [-ATR] vowel and additionally heads a [-ATR] span that it shares with the [-ATR] root vowel. This violates both the constraint for dominance FaithfulnessHEadSpan(+ATR) and the more general faithfulness constraint FaithfulnessHeadSpan(ATR), but crucially both are ranked below the markedness constraint on span heads. Note that the faithful candidate (18-a) also violates the markedness constraint on span heads *SpanHEAD $(\mathrm{e}, \mathrm{o})(+\operatorname{ATR})$, since $[\mathrm{e}]$ heads a simple span. This foreshadows that this constraint will become essential in banning underived [+ATR] mid vowels.

|  | *SH(eo) | SH-R | FHS(+A) | *AS | F-HS |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | *! |  |  | * |  |
|  |  |  | * |  | ** |
|  | *! |  |  |  | * |

Since a simple span must be headed by its only segment, *SpanHEAD(e,o)(+ATR) becomes active in isolated roots. Due to Richness of the Base, a $[+$ ATR $]$ mid vowel could also occur in monomorphemic forms in the input. It is necessarily the head of its own span because no other vowels are available as heads, as shown in the input representation in (19). Therefore the faithful candidate (19-a) can never become optimal. Instead, the [+ATR] mid vowel is neutralized to a [-ATR] mid vowel and candidate (19-b) becomes optimal. SpanHEAd $(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$ thus keeps $[+\mathrm{ATR}]$ mid vowels from both triggering harmony and occurring in underived environments.

No [+ATR] mid vowels in isolated roots because of head markedness

|  | SH(e,o) | SH-R | F-HS(+A) | *AS | FHS |
| :---: | :---: | :---: | :---: | :---: | :---: |
| a. | *! |  |  |  |  |
|  |  |  | * |  | * |

In sum, we have seen that in Assamese [e,o] cannot occur as triggers of regressive [+ATR] harmony. They also fail to show up in underived environments. Whereas a simplistic OT
analysis fails to account for this pattern, the representational power of Headed Spans Theory allows for a simple solution. The markedness constraint on heads $\operatorname{SpANHEAD}(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$ bans [ + ATR] mid vowels from heading a feature span. This explains both their absence in underived environments and as triggers of vowel harmony, since they would have to head a $[+\mathrm{ATR}]$ span in both contexts.

### 4.3 No Short Mid Triggers in Päri

Päri (Nilotic; South Sudan) has a system of root controlled [ATR] harmony (Andersen 1989). Vowel harmony is not dominant-recessive, i.e. it can be triggered by vowels with either $[ \pm$ ATR $]$ value and similarly result in $[+$ ATR $]$ vowels and $[-A T R]$ vowels. The [ $\pm$ ATR] value of a word is always determined by the value of the root vowel. The short vowel inventory of Päri is given in (20). Arrows indicate changes due to vowel harmony and triggers are underlined. As can be seen, low vowels do not undergo harmony. More crucially, mid $[+A T R]$ vowels cannot trigger harmony. Note that even though $/ \varepsilon /$ and $/ v /$ do not usually trigger harmony, I will focus on $[\mathrm{e}]$ and $[\mathrm{o}]$ here.
(20) Päri short vowel inventory


As shown in (21), vowel harmony applies both to prefixes and suffixes. Suffixes and prefixes usually include only one vowel each in all forms (Andersen 1989:1). In (21-a), the [-ATR] root vowel $/ \mathrm{I} /$ triggers a [-ATR] vowel in the suffix $-/ \mathrm{I} /$, which yields $-/ \mathrm{I} /$ in the harmonic surface form [wìnfí], whereas the [+ATR] vowel /i/ in the root /rìgg/ results in a [+ATR] vowel [í] in the same suffix in (21-b). (21-c) and (21-d) give analogous examples for prefixes.

Root controlled harmony
a. /wìnf/-/Í/ $\rightarrow$ [wìnfí] 'your bird'
b. /rìng/-/Í/ $\rightarrow$ rìingí] 'your meat'
c. /Í/-/dò:k/ $\rightarrow$ [ídò:k] 'you return'
d. /í/-/po:d/-/Ô/ $\rightarrow$ [ípo:dô] 'you beat (ANTIP)'

Usually, short /e,o/ cannot trigger harmony and do not occur in underived environments. A form like /á/-**/dòn/-/é/ in (22-a) would be impossible, since there are no roots with short $[+$ ATR $]$ mid vowels. They can, however, occur in certain morphological contexts, e.g. as the result of shortening mutation ( $=\mathrm{SM}$ ) in(22-b). In these contexts they can trigger harmony in the same way that other vowels trigger harmony. Note that the root / dòm/in (22-b) underlyingly includes a long vowel. On the surface, however, the vowel becomes short and triggers the [+ATR] variant -[é] of the suffix.
(22) Only derived harmony triggers
a. /á/-**/dòy/-/É/
b. /á/-/dò: $\mathfrak{y} /-/ E$ É/+SM $\rightarrow$ [ádòngé] 'he pushed it'

These data are again problematic for a simple parallel OT analysis. I still assume for the sake of simplicity that root-control and bidirectionality are both covered by a general Harmony (ATR) constraint. In (23), this constraint is violated by both the faithful candidate (23-a) and the expected candidate (23-c), since in the former no harmony applies and in the latter harmony applies in the wrong direction, i.e. from the suffix to the root. Therefore, the optimal candidate (23-b) allows for a short [+ATR] mid vowel as a trigger, contrary to the empirical facts. Note that the ranking Harmony(ATR) $\gg$ ${ }^{*} \mathrm{e}, \mathrm{o}$ is needed independently, since $[\mathrm{e}]$ and $[\mathrm{o}]$ need to be able to show up in affixes as the result of vowel harmony.

Failure of simplistic SPOT analysis

| I: /á/-/dò̀ /-/と́/ | Harmony (ATR) | ${ }^{*} \mathrm{e}, \mathrm{o}$ | Ident(ATR) |
| :---: | :---: | :---: | :---: |
| a. /á/-/dò̀y/-/غ́/ | *! |  |  |
| -b. /á/-/dò̀y/-/é/ |  | * | * |
|  | *! |  | * |

I will again show that this shortcoming can be resolved if autosegmental feature spans are headed and the representation of vowel harmony is therefore asymmetric. In a headed span analysis only one simple restriction on the head of an $[ \pm A T R]$ feature span can derive the inability of $[\mathrm{e}]$ and $[\mathrm{o}]$ to trigger harmony and their absence in underived environments. I will also show that the analysis predicts that $[\mathrm{e}]$ and $[\mathrm{o}]$ gain the ability to trigger harmony
if they surface for independent reasons in a triggering position, in this case in a root with length mutation. The constraint in (24) used in this analysis is virtually identical to the one used in the analysis of Assamese, except that it refers to short vowels, whereas length is not contrastive in Assamese.
*SpanHEAD (e,o) (+ATR)
Count one violation for every [ + ATR] span that is headed by a short mid vowel.

I follow Trommer (2011) in analyzing root control as involving stress on the root vowel. This means that root-control can be encoded in a constraint that requires the head of a span to be in the root/stressed. Again, *AdjacentSpans(ATR) triggers harmony and FaithfulnessHeadSpan(ATR) serves as a faithfulness constraint. Additionally, some constraint is needed that enforces mutation in the appropriate context, I will use a simplified constraint ShorteningMutation here and a corresponding general faithfulness constraint, without claiming that this provides a satisfactory analysis of mutation in Päri.
(25) Further constraints on Päri vowel harmony
a. SpÁNHÉÁd(ATR) (=ŚH́)

Count one violation for every $[ \pm$ ATR $]$ span that is not headed by a segment in the stressed syllable.
b. *AdJacentSpans(ATR) $(=*$ AS $)$

Count one violation for every each pair of adjacent [ $\pm \mathrm{ATR}]$ spans.
c. FaithfulnessHeadSpan(ATR) $(=\mathrm{FHS}(\mathrm{A}))$

Count one violation for a segment that is the head of a $[+ \pm$ ATR $]-S p a n ~ i n ~$ the input but not in the output.
d. ShorteningMutation (=SM)

Count one violation for any long root vowel in the shortening mutation context.
e. Faithfulness(Length) $(=\mathrm{F}(\mathrm{L}))$

Count one violation for any long root vowel in the shortening mutation context.

Short [+ATR] mid vowels are allowed in affixes if they are derived by vowel harmony. This can be analyzed using only the constraints for general harmony, as shown in the tableau in (26). SpÁnHÉÁd(ATR) is violated by any span headed by an affix vowel. This is true for the simple span of the faithful candidate in (26-a) as well as candidate ( $26-\mathrm{d}$ ) where the affix vowel spreads its [ATR] feature onto the root vowel. Note that the latter could serve as a strategy for avoiding short [ + ATR] mid vowels, but this is not necessary, since
the optimal candidate (26-b) does include such a vowel. This, however, does not violate the $* \operatorname{SpanHEAD}(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$, since the affix vowel is included in the $[+\mathrm{ATR}]$ span as a non-head. Including the affix vowel as a head of the $[+$ ATR] span as in candidate (26-c) creates an additional violation of this constraint, but this is rendered irrelevant by the fatal violation of SPÁnHÉÁd(ATR) due to the affix vowel serving as the head.

Derived short [ +ATR ] mid vowels as non-heads in Päri

|  | ŚH́ | *AS | *SH(e,o) | FHS |
| :---: | :---: | :---: | :---: | :---: |
|  | *! | * |  |  |
|  |  |  |  | * |
|  | *! |  | * | ** |
|  | *! |  |  | * |

Short mid vowels are banned from roots simply because of the *SpanHEAD (e,o)(+ATR) constraint. As seen in (27-c), any short [+ATR] mid vowel in a root is banned from surfacing because it is necessarily the head of a $[+$ ATR $]$ span. This is also true for simple spans, as in the faithful candidate (27-b). I will assume that this constraint violation is repaired by altering the length specification of the vowel in question, as in (27-b). The faithfulness constraint for length has to be ranked lower anyway in order to allow shortening mutation. Since the head vowel is no longer a short vowel, the constraint *SpanHEAd $(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$ is satisfied.
(27)

No short [+ATR] mid triggers in Päri vowel harmony

|  | ŚH́ | *AS | *SH(e,o) | FHS | F(L) |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | *! | * | * |  |  |
|  |  |  |  | * | * |
|  |  |  | *! | * |  |

A higher ranked constraint can overrule the *SpanHEAD $(\mathrm{e}, \mathrm{o})(+\mathrm{ATR})$ constraint and short [ + ATR] mid vowels can regain their ability to trigger vowel harmony. This is the case in Shortening Mutation contexts, as shown in the tableau in (28). The higher ranked constraint ShorteningMutation excludes candidate (28-b) with a lengthened vowel that becomes optimal in other contexts. Instead, the short mid vowel can serve as the head of a $[+\mathrm{ATR}]$ span in this exceptional context, as can be seen in candidate (28-c).
(28) Derived short [+ATR] mid vowels trigger vowel harmony in Päri

|  | SM | *AS | *SH(e,o) | FHS | F (L) |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | *! | * | * |  |  |
|  | *! |  |  | * |  |
|  |  |  | * | * | * |

All in all, the Päri data have shown to be empirically similar to the Assamese data.

In both cases, $[+A T R]$ mid vowels can neither trigger harmony nor occur in underived environments. They solely occur as the result of vowel harmony. Päri however has shown that this generalization needs not to be exceptionless. In a morphologically derived environment, mid vowels can regain their ability to trigger vowel harmony. This is expected in OT, where higher ranked constraints can derive exceptions to more general patterns.

### 4.4 Output Drivenness and Previous Approaches

In this section, I will compare three approaches that have been used inside and outside OT phonology to derive the asymmetric non-trigger pattern. I will argue that constraints on underlying representations used in pre-OT autosegmental phonology are empirically adequate but obviously incompatible with Richness of the Base. Crucially, the root strata approach used in Trommer (2011) is conceptually and empirically indistinguishable from such an approach and I conclude that it does not allow us to uphold the Richness of the Base principle. A stricter criterion will be introduced in this section: Output Drivenness. I will show that including complex faithfulness constraints as used by Mahanta (2008) for Assamese does not fulfill this criterion and allows us to effectively exclude inputs from mapping them to expected outputs. This is in stark contrast to the headed spans approach developed in the previous section, which neither uses complex faithfulness nor restrictions on the input.

Tesar (2013) characterizes Optimality Theory as a surface-oriented framework. This means that differences between the input and the output can only be explained by restrictions on surface forms. This is crucially related to Richness of the Base, as I will show in this subsection. It is a stronger notion, however, since not only are language-specific restrictions on the input banned, output restrictions are the only mechanism that can lead to unfaithful outputs. Tesar further formalizes this notion as Output Drivenness in (29). This can be paraphrased as stating that any mapping of an input to an output also implies mapping any more similar input to the same output.

## (29) Output Drivenness

A phonological map is output-driven if, for any mapping from an input to an output, any other input that has greater similarity to the output also maps to the same output.

Tesar further shows that only faithfulness constraints can be problematic for this property in Optimality Theory. He provides a justification for why certain classes of complex faithfulness constraints are not output-driven preserving. These constraints are problematic for Richness of the Base, since they allow certain inputs to not reach the surface as expected.

In the simplest case this means that an output, which would also be a grammatical output, does not map to itself. Such a map is made possible by certain classes of constraints, e.g. by complex faithfulness of the type proposed in Mahanta (2008).

More specifically, two constraint classes are relevant here: disjunctive faithfulness and faithfulness conditioned by an independent context. Both of these predict non-outputdriven patterns. Disjunctive faithfulness, i.e. a faithfulness constraint that is violated by either a change in a feature F or in another Feature G, predicts phonologically derived environment effects. The problem is that it does not distinguish between candidates that violate both constraints or just one of the two. This means that a fully faithful candidate can be distinguished from any less faithful candidate, but any two unfaithful candidates are treated the same. Tesar terms this behavior 'distinction only at greater similarity'. This means that a less similar input can be mapped to an output that is excluded for a more similar output.

Tesar uses vowel raising as a more concrete example. Disjunctive faithfulness allows for a derived environment pattern ${ }^{11}$ in vowel raising, as shown in (30) and (31). The low vowel /a/ maps to [ i$]$ in candidate (30-b). The input /e/, which is more similar to [i], however, maps to itself in candidate (31-a). This output is not possible for input /a/, as candidate (30-c) cannot become optimal. This is only possible due to the disjunctive faithfulness constraint. It allows the low ranked markedness constraint *[-high] to become active for input /a/ in (30), even though candidate (30-c) is more faithful to the input compared to candidate $(30-\mathrm{b})$. Note that this concerns the features mentioned in the disjunctive constraint. This can be interpreted as banning the input /a/ from becoming expected output [e] by mapping it to a less faithful candidate.
/a/ neutralizes to less similar [i] instead of more similar [e] (Tesar 2013:124)

| I: | a | *[+low] | IDENT(low) $\vee$ IDENT(high) | *[-high] |
| ---: | ---: | :---: | :---: | :---: |
| a. | a | $*!$ |  | $*$ |
| b. | i |  | $*$ |  |
| c. | e |  | $*$ | $*!$ |

[^51][e] is allowed if it results from an underlying /e/ (Tesar 2013

| I: | e | [+low] | IDENT(low) $\vee$ IDENT(high) | $*[-$ high $]$ |
| ---: | ---: | :---: | :---: | :---: | :---: |
| a. | e |  |  | $*$ |
| b. | a | $*!$ | $*$ | $*$ |
| c. | i |  | $*$ |  |

A similar argument can be constructed for constraints with input restrictions. Tesar describes these constraints as amounting to a local conjunction of a constraint that is violated by an input context with a standard faithfulness constraint. These constraints predict chain shift patterns. Using vowel raising again as an illustration in the tableaux in (32) and (33), we find that a non-output-driven pattern results. This is because the constraint only makes a distinction at lesser similarity. Violations of Ident(high) are counted only for the input low vowel /a/ in (32), keeping it from becoming the maximally unmarked [i] in candidate (32-c) and mapping it to [e] in candidate (32-b) instead. [i] in candidate $(33-\mathrm{c})$ is however a possible output if the input is [e]. This pattern is non-output-driven, since it is more similar to output [e] than /a/, yet it maps to less similar [i]. One interpretation of this pattern is that the input /a/ is excluded from mapping to the expected unmarked [i].

Input /a/ maps to [e] instead of less marked [i]

| I: | a | $*+$ low | ${ }^{*}+$ low $_{\text {Input }} \& \operatorname{IDENT}(\mathrm{hi})$ | $*[$-high $]$ |
| ---: | ---: | :---: | :---: | :---: |
| a. | a | $*!$ |  | $*$ |
| b. | e |  |  | $*$ |
| c. | i |  | $*!$ |  |

[i] is allowed if it derives from underlying /e/

| I: | e | $*[+$ low $]$ | $*+$ hi $_{\text {Input }}$ \& IdENT(hi) | $*[$-high $]$ |
| ---: | ---: | :---: | :---: | :---: |
| a. | e |  |  | $*!$ |
| b. | a | $*!$ |  | $*$ |
| c. | i |  |  |  |

All in all, these constraint types undermine the Output Drivenness of Optimality Theory. A more concrete discussion of Mahanta's analysis of Assamese is given in 4.4.3. It should be noted that the faithfulness constraint in Headed Spans Theory crucially is not analyzable as the conjunction of a faithfulness constraint with a markedness constraint or a faithfulness constraint as I argue in section 4.4.4.

### 4.4.1 Constraints on Underlying Representations Ban Non-Triggers

Clements (1984) argues against a spiritual predecessor of Optimality Theory called WSC (Word Structure Conditions), which is based on surface-true structure-preserving conditions associated with specific repair rules as proposed by Stewart (1983) for Akan. One main difference between the two analyses is the treatment of the [+ATR] counterpart of the low vowel /a/. Clements (1984) argues in favor of a Morpheme Structure Constraint mainly for conceptual reasons. The basic empirical pattern is relatively simple. Low vowels become [+ATR] before high [+ATR] vowels. Elsewhere, they are [-ATR]. This is an instance of the asymmetric non-trigger pattern, since [-ATR] low vowels cannot occur in a root without a following high vowel in this root-controlled harmony system. ${ }^{12}$
(34) Distribution of [a] and [ə] (Clements 1985:325,326)
a. bisa, ${ }^{* *}$ bisə
'to ask'
b. kasa, **kəsə
'to speak'
c. kəri, *kari
'to weigh'
d. w-ə-di, *w-a-di
'he has eaten'

Clements (1985) proposes a constraint on underlying representations for segments, given in (35). This condition implies a [-ATR] value for all low vowels. In other words, it bans [ + ATR] low vowels from underlying representations.
(35) CoUR for low vowels in Akan

$$
[+ \text { syllabic },+ \text { low }] \Rightarrow[-\mathrm{ATR}]
$$

Additionally, a late raising rule is postulated (cf. also Kiparsky 1985:121). This rule raises [-ATR] low vowel to the [+ATR] vowel / / before high vowels. This accounts for the derived occurrences of the [+ATR] counterpart of the low vowel. These two assumptions easily work for the attested distribution. If the input includes a /ə/, e.g. **/kəsə/, it is

[^52]excluded by the CoUR in (35). Crucially, even though this segment is excluded in the input, it can be derived by the raising rule. This leads to a situation where forms such as /kəri/ and /ədi/ are excluded as an input but licit as an output, cf. (36). This is a case of counterfeeding interaction, since the application of raising comes to late to be affected by the ban on underlying $/ \mathrm{\rho} /$.
(36) Derivation of the distribution of low vowels in Akan using CoUR and a raising rule

| Input: | bisa | bisə | kasa | kəsə | kari | kəri | w-a-di | w-ə-di |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| CoUR: | n.a. | - | n.a. | - | n.a. | - | n.a. | - |
| Raising | n.a. | - | n.a. | - | kəri | - | wədi | - |
| Output: | bisa | - | kasa | - | kəri | - | wədi | - |

According to Clements (1984), one major advantage of an approach based on constraints on underlying representations is the conservation of the notion of a phoneme. Only contrastive segments can occur underlyingly. Interestingly, this very same requirement has led to the Duplication Problem (cf. chapter 1), which has received a solution in the Richness of the Base principle in OT. One might therefore argue that the main argument in favor of an CoUR based approach is rendered moot.

The second argument was already foreshadowed in the introduction of this chapter. Clements (1984) argues that a surface oriented based approach is more complex due to the asymmetric non-trigger problem. Under the assumptions brought forward in (Stewart 1983), two conditions are needed in order to account for the distribution of [a] and [ $\partial$ ], a bidirectional one for roots and a unidirectional one for prefixes. Two assumptions are crucial here. For one, Stewart assumes that only high vowels trigger harmony, contrary to the empirical data brought forward by Clements (1984, 1985). This means that low vowels are regular undergoers for vowel harmony. Secondly, all affix vowels are underlyingly [-ATR]. Therefore assimilation is birectional for root vowels (since both / $/$ / and $/ \mathrm{a} /$ occur in roots) but unidirectional in affixes, where underlyingly only /a/ is present. The second assumptions is meant to capture the facts about root control, but it clearly instantiates a language-specific constraint on underlying representations. If this assumption is abandoned, Stewart's approach gets closer to adhering to the modern concept of Richness of the Base. If additionally root control is brought about by other means, the need for two different conditions vanishes and one bidirectional condition suffices.

I do not wish to argue against the empirical adequacy of this approach here. It is obvious, however, that such an approach is inconsistent with the assumption of Richness of the Base, since language specific constraints on the input are employed. Nevertheless, I would like to point out that this approach has been largely remained unquestioned in
the literature on vowel harmony and is still influential. Following Kiparsky (1985), the distinction between structure-preserving processes that change phonemes into phonemes and structure-changing processes that create new allophones has persisted in discussions of vowel harmony. Recent empirical (Rolle et al. 2020) and theoretical studies (van der Hulst 2018) have kept the distinction, using terms like allophonic, post-lexical or postcyclic vowel harmony. Language specific restrictions on inventories or underlying feature specifications have remained a viable alternative, in Optimality Theory approaches (Casali 2003) and in other representational theories (van der Hulst 2018). Casali (2003:365) proposes that there should be a language-specific constraint on underlying specification of [ $\pm \mathrm{ATR}$ ] values in order to derive certain typological facts about vowel harmony and related phenomena. In some languages, only [+ATR] can ever be specified and there is no underlying [-ATR] feature. Other languages instantiate the mirror image with underlyingly specified [-ATR] and no [+ATR]. Since this is a restriction on underlying representations that is not universal, this would be a case of an OT system without Richness of the Base.

### 4.4.2 Root Strata Neutralize Non-Triggers

Trommer (2011) proposes an analysis of Päri vowel harmony that makes crucial use of a root stratum in a stratal OT framework. Phonological computation at the root stratum takes place before any other phonological and morphological operation. It includes a language specific constraint ranking and its domain only consists of monomorphemic phonological objects. I will argue that this makes it impossible to distinguish root strata from Morpheme Structure Constraints both empirically and conceptually.

The first part of the analysis is very similar to Clements (1985) analysis of Akan. Since harmony is root controlled, short [+ATR] mid vowels are banned from occurring in root morphemes by a high ranked constraint at the root level. Non-low affix vowels become underspecified for $[ \pm$ ATR $]$. At the later stem and word level, a constraint requiring vowel harmony is ranked higher and short [ + ATR] mid vowels can emerge as the result of vowel harmony in affixes. Additionally, mutation at the word level can derive short mid [+ATR] vowels in roots.

Looking at the table in (37), we see the outcomes of each stratum for a hypothetical short [ o ] root, a long [ $\mathrm{o}:$ ] root and a long [ $\mathrm{o}:$ ] plus shortening mutation. At the root stratum, the difference between long and short [o] roots is neutralized. From this point on they are indistinguishable. Additionally, the affix vowel /e/ becomes underspecified for ATR, indicated with a capital letter $/ \mathrm{E} /$. Note that at this point, morphological concatenation has not taken place yet, indicated by the hyphens between exponents. At the stem stratum, segmental affixes are attached and vowel harmony applies. Shortening mutation applies afterwards at the word stratum, thereby counterfeeding neutralization at the root stratum.

| Input: | á-dò̀y-é | á-dòòı-é | á-dò̀ıı́a + SM |
| :---: | :---: | :---: | :---: |
| Root Stratum: | á-dò̀ı!-É | á-dò̀: y -É | á-dò̀ y -É+SM |
| Stem Stratum: | ádò: ${ }^{\text {áe }}$ | ádò: gé $^{\text {a }}$ | ádò: éé + SM $^{\text {a }}$ |
| Word Stratum: | ádò: ${ }^{\text {dé }}$ | ádò: gé $^{\text {a }}$ | ádòyé |
| Output: | ádò: $\mathrm{m}_{\text {é }}$ | ádò: $\mathrm{I}_{\text {é }}$ | ádòyé |

In chapter 1, we have seen that Morpheme Structure Constraints can be conceptually defined by three properties, language specificity, their level and their domain. We have also noted that MSCs have taken many different forms in different theories, including positive schemata, negative constraints, rules and conditions. I see no reasons, why a language-specific ranking of OT constraints should not be included here, even if it is a more complex grammatical building block. Root strata in stratal OT thus fulfill all three defining conditions of an MSC. In addition to the language-specific ranking, they apply before any other morphological and phonological operation and take simple morphemes as their domain for phonological computation. Conceptually, root strata thus become indistinguishable from Morpheme Structure Constraints. I concur that they therefore violate the Richness of the Base principle. They impose language-specific restrictions on the input of the stem level, the first level that includes polymorphemic domains and phonological alternations.

## Morpheme Structure Constraints

a. MSCs apply before any other morphological or phonological operations.
b. MSCs apply to monomorphemic forms.
c. MSCs are language-specific.

Empirically, we have seen that root strata can derive the asymmetric non-trigger pattern in a manner analogous to Morpheme Structure Constraints by serial interaction. The root level ban on short [+ATR] mid vowels bleeds the application of [+ATR] vowel harmony at the stem/word level. Additionally, the application of vowel harmony counterfeeds the deletion of ATR specifications on affixes. This pattern is thus modeled using the same basic logic. I conclude that it is very likely that root strata make the same empirical predictions that we find for MSCs.

### 4.4.3 Complex Faithfulness for Triggers Only

For a simple Optimality Theory approach the challenge posed by Assamese vowel harmony is that both dominance and directionality enforce /e,o/ as triggers. Mahanta (2008) solves this problem by separating dominance in a dedicated markedness constraint and restricting
it to high vowels: ${ }_{\mathrm{I}, \mathrm{v}}$. This however creates a problem with directionality, since it would predict that [-ATR] high vowels would assimilate to [+ATR] vowels in any direction. ${ }^{13}$ Mahanta's solution for this new problem is a faithfulness constraint that protects [ $\pm \mathrm{ATR}$ ] values only for high vowels. This leads to reversed rankings of faithfulness and markedness constraints for high vs. mid vowels, cf. (39). [ $\pm \mathrm{ATR}]$ contrast in high vowels is preserved. In mid vowels, however, it is neutralized to the unmarked value, unless vowel harmony demands something different.

$$
\begin{align*}
& \text { Mahanta's final ranking }  \tag{39}\\
& *[-\mathrm{ATR}][+\mathrm{ATR}] \gg \mathrm{ID}(+ \text { high })(\mathrm{ATR}) \gg *_{\mathrm{I}, v} \gg *_{\mathrm{e}, \mathrm{o}} \gg \mathrm{ID}(\mathrm{ATR})
\end{align*}
$$

The tableau for mid vowels is thus minimally different. Since the high ranked harmony constraint does not include directionality and dominance of $[+\mathrm{ATR}]$ anymore, the candidate with progressive assimilation to [-ATR] can become optimal. In (40), the missing violation in candidate ( $40-\mathrm{c}$ ) is marked with a circle.

Harmony without dominance predicts the correct output

| I: | /kər-o/ | $\left.*^{[-A T R}\right][+$ ATR $]$ | Id(+high)(ATR) | $*_{\mathrm{I}, v}$ | $*_{\mathrm{e}, \mathrm{o}}$ | Id(ATR) |
| ---: | :--- | :---: | :---: | :---: | :---: | :---: |
| a. | kəro | $*!$ |  |  |  |  |
| b. | koro |  |  |  | $*!$ | $*$ |
| c. | kərə | $\bigcirc$ |  |  |  | $*$ |

Without the complex faithfulness constraint, the ranking in Mahanta (2008) would predict bidirectional assimilation of high values to [+ATR] vowels, due to the markedness constraint $*_{I, v}$ that derives [+ATR] dominance. The faithfulness constraint ID(+high)(ATR) saves the faithful candidate in (41-a) by excluding any change that is not due to directional vowel harmony.

Complex faithfulness saves triggers from bidirectionality

| I: /phur-v/ | *[-ATR][+ATR] | Id (+high)(ATR) | $*_{1, v}$ | $*_{\text {e,o }}$ | Id (ATR) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| a. $\mathrm{p}^{\mathrm{h}}$ urv |  |  | * |  |  |
| b. $\mathrm{p}^{\mathrm{h}}$ uru |  | *! |  |  | * |
| c. $\mathrm{p}^{\mathrm{h}}$ ชrv |  | *! | ** |  | * |

Let us look closer at the complex faithfulness constraint. Mahanta (2008) describes its

[^53]satisfaction condition as a logical conjunction of Ident constraints on the features [+high] and $[ \pm \mathrm{ATR}]$. This implies that the violation condition is the local disjunction of the two constraints. Since this is not immediately clear from the description, I give the constraint disjunction in (42).
$\operatorname{IdENT}(+$ hi) $\vee \operatorname{IdENT}(A T R)$
Count one violation for any vowel that is
a. [+high] in the input and [-high] in the output or
b. $[\alpha \mathrm{ATR}]$ in the input and $[-\alpha \mathrm{ATR}]$ in the output.

A correspondence-theoretic disjunctive faithfulness constraint undermines the Output Drivenness of OT (cf. Tesar 2013). It allows to predict phonologically derived environment effects. In (43), input $/ v /$ can change both $[ \pm$ ATR $]$ and $[ \pm$ high $]$ features resulting in $[0]$ in candidate (43-b). Under the ranking shown, the more faithful candidate (43-c) [u] cannot emerge as the winner due to a low ranked markedness constraint. This is despite the fact that $[\mathrm{u}]$ is a possible output itself, as shown in (44) for candidate (43-a). The reason is that the constraint disjunction does not distinguish between violating faithfulness once or twice. A derived $[u]$ is not allowed and the vowel has to change a further feature to become licit. An underlying [u] however surfaces faithfully. This excludes the input /v/ which would otherwise be expected to map to $[\mathrm{u}]$ - by neutralizing it with another, less similar vowel. ${ }^{14}$

Input /v/ is neutralized to a less similar vowel $[\mathrm{o}]$ instead of the more similar $[\mathrm{u}]$

| I: v | *[-ATR] | $\operatorname{IDENT}(+h \mathrm{i}) \vee \operatorname{IDENT}(\mathrm{ATR})$ | $*[+$ high $]$ |
| :---: | :---: | :---: | :---: |
| a. $\quad$ | *! |  | * |
| amb. O |  | * |  |
| c. u |  | * | *! |

[u] is allowed if it results from an underlying $/ \mathrm{u} /$

| I: | u | $*[-$ ATR $]$ | IDENT(+hi) $\vee \operatorname{IdENT}($ ATR $)$ | $*[+$ high $]$ |
| ---: | ---: | :---: | :---: | :---: |
| a. | u |  |  | $*$ |
| b. | v | $*!$ | $*$ | $*$ |
| c. | o |  | $*$ |  |

There is no such effect in the analysis in Mahanta (2008), mainly due to the high ranking IDENT(high) faithfulness constraint that renders the first disjunct of the constraint inactive.

[^54]Additionally, such a constraint would face empirical challenges, e.g. the input /koro/ would be mapped to an unattested output [koro], since the input vowel sequence does not violate *[-ATR][+ATR]. Under a more gracious interpretation, IDENT(+high)(ATR) would be an IDENT(ATR) constraint that simply applied only to input high vowels. This is supported by another definition provided by Mahanta (2008:118). Such a constraint would thus be a faithfulness constraint with an input restriction. As shown by Tesar (2013), faithfulness constraints conditioned by the input context show a different set of problematic properties. They predict chain-shift-like patterns, which are also non-output-driven. He notes that logically, such a constraint can be understood as a conjunction of two parts. The first part assigns a violation, every time the input condition is met, in this case every time a [+high] vowel shows up in the input. ${ }^{15}$ The second part is just a simple faithfulness constraint, in our case Ident(ATR). These two constraints are locally conjoined with the vowel as the domain.
$\left.*_{[+h i g h}\right]_{\text {Input }} \&_{V} \operatorname{IDENT}(A T R)$
Count one violation for every vowel that
a. is [+high] in the input and
b. changed its $[ \pm \mathrm{ATR}]$ value from input to output.

As shown in (46), input high vowels are now banned from changing their $[ \pm \mathrm{ATR}]$ value, even if this would reduce markedness. The optimal candidate (46-b) wins over the less marked candidate ( $46-\mathrm{c}$ ) due to the complex faithfulness constraint. The same output vowel $[\mathrm{o}]$, with a change in $[ \pm \mathrm{ATR}]$, is allowed on the surface if it results from an input $/ \nu / .{ }^{16}$ Effectively, input $/ v /$ is excluded from mapping to the expected less marked output [ o ] due to the complex faithfulness constraint. The resulting pattern is a chain shift: $v \rightarrow$ っ, 〕 $\rightarrow$ о.
Input $/ v /$ maps to $[0]$ instead of less marked [o]

| I: | $v$ | $*[+$ high $]$ | $*[+ \text { high }]_{\text {Input }} \&_{\text {V }}$ IdENT(ATR) | $*[-$ ATR $]$ |
| ---: | ---: | :---: | :---: | :---: |
| a. $v$ | $*!$ |  | $*$ |  |
| b. | $v$ |  |  | $*$ |
| c. | o |  | $*!$ |  |

[^55][o] is allowed if it derives from underlying /o/

| $\mathrm{I}:$ | $\supset$ | $*[+$ high $]$ | $*[+ \text { high }]_{\text {Input }} \&_{\mathrm{V}}$ IDENT(ATR) | $*[$-ATR $]$ |
| :---: | :---: | :---: | :---: | :---: |
| a. | $\supset$ |  |  | $*!$ |
| b. | $\mho$ | $*!$ |  | $*$ |
| c. | o |  |  |  |

In sum, the approach in Mahanta (2008) is empirically adequate if a faithfulness constraint with an input context is assumed. It derives the Assamese asymmetric non-trigger pattern. The device used however is problematic, since it undermines the Output Drivenness of Optimality Theory. It allows us to exclude unwanted input-output mappings by neutralizing the input with a different output from the one that would be expected from markedness and simple faithfulness considerations. If constraint disjunction is assumed, phonologically derived environment effects are expected and if a context-restricted faithfulness constraint is assumed, chain shifts are predicted.

The results are ambiguous. One could take the empirical predictions as an advantage if such data exist and cannot be derived in any other way. On the other hand, one could also focus on the conceptual disadvantage, since OT loses its Output Drivenness, even if the input is not directly restricted in a language specific way. The lack of Output Drivenness might lead to learnability problems (cf. Tesar 2013). Either way, direct constraints on the input are not needed and Richness of the Base can be maintained.

### 4.4.4 Faithfulness in Headed Spans Theory

The faithfulness constraint in Headed Spans Theory needs special comment. McCarthy (2004) assumes that it refers to head status and feature content at the same time. This property crucially allows Span Theory to exclude standard sour grapes patterns (Finley 2008:41). It also requires us to make explicit assumptions on the autosegmental structures in the input. In many autosegmental analysis it is more or less implicitly assumed that underlyingly there is maximally a one-to-one relation between features and segments (cf. McCarthy 2004). This means that the structure in (48-a) is a possible input, but (48-b) is not. Under such an assumption, the interpretation of the constraint is straightforward in the input, every segment is a head. In the output, the faithfulness constraint is violated for any dependent segment and for any segment that changes its feature value. Note that Richness of the Base can still be maintained. The input restriction against non-minimal spans is not language-specific. Therefore, it does not create any differences between the input set in different languages.
(48) Simple and complex headed spans


Under a different interpretation of the faithfulness constraint, it is also compatible with one-to-many mappings in the input. This means that any segment that is linked to a feature in the input, needs to head a feature span in the output (cf. Hudu 2010). Any segment that is linked to a feature in the input and does not head a feature span in the output would count as a violation of such a constraint. This also means that splitting up a feature span would not produce any faithfulness violations. An input structure (48-b) could be mapped to an output structure (48-a) without any faithfulness violations, only increasing markedness constraint violations.

Another aspect of the faithfulness constraint is its complex nature. O'Keefe (2007) and Akinlabi (2009) mention in a footnote that this faithfulness constraint is really a conjunction of a markedness constraint against non-heads of a certain feature value and a faithfulness constraint to a certain feature value. Similarly, Orgun \& Sprouse (2008) propose to split up the constraint into two faithfulness constraints, one for head status and one for a feature value, i.e. a usual Ident constraint. I would like to argue that such a conception is not desirable for two reasons. First, if we assume that the constraint is a conjunction of a constraint against non-heads of a feature value $[\alpha \mathrm{F}]$ and an IDENT constraint to the same feature value, there is no reason why there should not be a conjoined constraint for different features, e.g. a constraint that requires segments to be faithful to their [ATR] value and to head a [+high] span. Alternatively, if we interpret the conjoined constraint as only including the Ident constraint for a certain feature value, what would the other constraint be? It would need to be a constraint that preserves the head status of a segment. This would mean that the simple constraint would be satisfied, if a segment that is the head of a span of the feature F in the input, becomes the head of a feature G in the output. This seems to make jarringly wrong empirical predictions. Secondly, Tesar (2013) has shown that constraint conjunction with faithfulness constraints leads to a non-output-driven pattern. In an surface-oriented theory like Optimality Theory, such constraints should be avoided. This is true for both faithfulness-faithfulness conjunction and markedness-faithfulness conjunctions. They could covertly reintroduce restrictions on the input.

I propose that the notion of headedness has no status independent of a span with certain feature value. There is no independent feature [ $\pm$ head] on a segment. Instead, there are three possible (non-)relations between a feature and a segment. In the first option the feature and the segment are not associated to each other. The second possible relation is that the segment is the head of a feature span. Finally, a segment can be a non-head of a feature span. These three relations form an unordered set and faithfulness constraints cannot refer to natural classes of these relations. We could expect that there is a constraint that requires a non-head of a feature span with feature value $\alpha \mathrm{F}$ to stay a non-head of the same feature if we allow for multiply linked features in the input, but no constraint can require faithfulness to general headedness or non-headedness independent of a span with a certain feature value.

### 4.5 Crosslinguistic Data

Trigger asymmetry patterns are found in other languages. In a survey on languages of the Macro-Sudan Belt in Africa, Rolle et al. (2020) find several languages with what they call derived-only patterns. In 18 languages from the Central Sudanic, Koman, Nilotic, and Nubian language families and from Bantoid languages of the Atlantic-Congo family [+ATR] mid vowels are asymmetric non-triggers. This means that in these languages, [+ATR] mid vowels do not occur in underived environments and cannot occur as triggers of harmony. Kakwa (Nilotic; Democratic Republic of Congo, South Sudan, Uganda) shows an instance of such a pattern. Even though [+ATR] non-low vowels are required to only cooccur with other [+ATR] vowels, a word solely consisting of [+ATR] mid vowels is impossible. The vowels thus occur as a result of vowel harmony but cannot trigger it.
(49) No isolated [+ATR] mid vowels in Kakwa (Onziga \& Gilley 2012:9)
a. lemi
'cause, obligation, right'
b. píré
'to fatten'
c. ${ }^{* *}$ leme
d. ${ }^{* *}$ pere

Similarly, in Avikam (Atlantic-Congo; Côte d'Ivoire) as well as in several varieties of Yoruba, Ijesa Yoruba and Ekiti-Irun-Ifaki Yoruba (Atlantic-Congo; Nigeria), [-ATR] high vowels only occur if derived by vowel harmony. This also precludes them from triggering ATR harmony. In Ekiti Yoruba, for example, [-ATR] high vowels cannot occur in root-final position, cf. (50-a) This is exactly the position where triggers of [-ATR] vowel harmony
occur. [-ATR] high vowels thus cannot trigger harmony. (50-d,e), shows that they can occur as the result of harmony if the root-final vowel is a non-high [-ATR] vowel. This pattern could alternatively be described as a positional markedness restriction, with high [-ATR] vowels not occurring in root-final position. In total, Rolle et al. (2020) find 23 derived-only patterns in their survey of 681 languages, which includes both languages with and without ATR-harmony.
(50) No high [-ATR] vowel harmony triggers in Ekiti Yoruba (Orie 2003)
a. bi
'ask'
b. èbi
'guilt'
c. ${ }^{* *}{ }_{\mathrm{b} I},{ }^{* *}{ }^{\text {è }}{ }_{\mathrm{b}}$
d. $\tau \mathrm{d} \varepsilon$
'brass'
e. vgbá
'calabash'

Casali (2003) adds to this another 14 languages where the [+ATR] counterpart of the low vowel /a/, only occurs 'allophonically'. This means that it only occurs when derived by vowel harmony but not in underived environments. It never triggers vowel harmony. This pattern is found in 14 languages from different subfamilies of the Atlantic-Congo family, i.e. Kwa, Gur and Bantoid but also in Surmic, Furan and Central Sudanic languages. We have seen an example of such a language in the previous sections with data from Akan. All in all, Casali (2003) finds 32 languages with a derived-non-trigger pattern in his sample of 110 languages.

In his survey on vowel harmony ${ }^{17}$ in the languages of the world, van der Hulst (2018) finds 15 languages - mainly Bantoid languages of the Atlantic-Congo family - whose vowel harmony he terms 'allophonic' or 'post-cyclic'. This again means that these vowels do not occur as triggers of vowel harmony or in underived contexts. Seven of these cases involve cases where [+ATR] mid vowels are asymmetric non-triggers. Additionally, in Northern Sotho (Atlantic-Congo, Botswana, South Africa), Sesotho (Atlantic-Congo; Lesotho, South Africa) and Setswana (Atlantic-Congo; Botswana, Namibia, South Africa, Zambia, Zimbabwe) a special series of raised high vowels can only occur as the result of raising [-ATR] high vowels before [+ATR] high vowels. This third series of high vowels is

[^56]distinct from the other two series. Two cases involve a subset of [-ATR] vowels being unable to trigger to trigger [-ATR] harmony, namely [ $\mathrm{I}, \varepsilon, ~\lrcorner]$ in Pasiego Spanish (Indo-European, Spain) ${ }^{18}$ and $[\varepsilon]$ in Kaba (Central Sudanic, Central African Republic, Chad).

Apart from [i] being unable to trigger raising harmony in Lhasa Tibetan (Sino-Tibetan, China), the most interesting case is Southwestern Khanty (Uralic, Russia) because this is a case of backness harmony, whereas all other cases reported refer to vowel height in some way or another. Here, the marked vowels $[y, u, æ]$ do not occur in word-initial syllables and vowel harmony is progressive. ${ }^{19}$ As shown in (51), [y, w, æ] can occur if derived by vowel harmony, but not as the sole vowel of a root. They are thus banned from triggering harmony, but this could be described as a positional restriction as well, similar to the Ekiti Yoruba data above. This hints at a reason for the absence of asymmetric non-triggers in many descriptions of vowel harmony. Since most cases of backness harmony are root-controlled and left-to-right, due to the suffixing nature of these language, the two descriptions cannot be told apart easily. Whereas dominant-recessive, bidirectional and non-root controlled directional harmony are abundant in height harmony systems, they are rare in systems where vowels harmonize for other features.
(51) Asymmetric non-triggers in Southwestern Khanty (Vaysman 2009:110,113)
a. uw-ə乏ux, ${ }^{* *}$ uw-
door-ABESS
b. lipt-yt, ${ }^{* *}$ lypt-
feed-nMLZ
c. nøməs-næ, **næməs-
mind-LOC

This generalization is also mirrored by the areal distribution of asymmetric non-triggers. The vast majority (50 languages) are found in Africa. Only a small number (6 languages) occur in Eurasia. Again, this can be linked to the abundance of bidirectional and dominantregressive vowel harmony patterns in Africa and their relatively low number in Eurasia.
(52) Areal distribution of asymmetric non-triggers

| Macro-Area | number of lgs. with asymmetric non-triggers |
| :--- | :--- |
| Africa | 50 |
| Eurasia | 6 |
| Total | 56 |

[^57]The language-family with the most languages (32 languages) with asymmetric non-triggers is Atlantic-Congo. This is hardly surprising, since it is the largest language family with frequent vowel harmony patterns ( 889 subfamilies, 1403 languages, Hammarström et al. 2018). Other notable African language families on the list are Central Sudanic and Nilotic, which are much smaller overall ( 64 and 55 languages, respectively) but rank second and third with 8 languages and 5 languages respectively. Language families from Eurasia include Indo-European (3 languages), Uralic (2 languages) and Sino-Tibetan (1 language). One text book example of vowel harmony languages is missing, namely the Turkic languages. This, again, might be due to the strict directionality and root-control in Turkic vowel harmony systems.

| Genealogical distribution of asymmetric non-triggers |  |
| :--- | :--- |
| Language Family | number of lgs. with asymmetric non-triggers |
| Atlantic-Congo | 32 |
| Central Sudanic | 8 |
| Nilotic | 5 |
| Indo-European | 3 |
| Koman | 2 |
| Uralic | 2 |
| Sino-Tibetan | 1 |
| Nubian | 1 |
| Furan | 1 |
| Surmic | 1 |
| Total | 56 |

These data show that the pattern of asymmetric triggers in vowel harmony is - even if not the default case - relatively common. It also seems that the same reoccurring vowel qualities are banned from triggering [ATR] vowel harmony. This is consistent with the findings in Archangeli \& Pulleyblank (1994) that low [+ATR] vowel and high [-ATR] vowel are more marked than their counterparts. This small scale study shows a consistent pattern, even though this might of course turn out to be a genealogical and/or areal effect. In $[ \pm \mathrm{ATR}]$ harmony systems, [e] and [ o$]$ are the most common asymmetric triggers (26 languages). Note that they sometimes cooccur with [a] (5 languages), which also occurs on its own (10 languages). All of these generally occur in a pattern that involve assimilation towards [+ATR] vowels. The relatively low frequency of [-ATR,+high] vowels as asymmetric non-triggers (6 languages) can be explained by the general scarcity of vowel harmony systems where the feature [-ATR] is dominant. Apart from the $\left[\begin{array}{l}1 \\ ,\end{array}, \underset{\sim}{c}\right]$ not triggering raising in closely related Northern Sotho, Sesotho, and Setswana (3 languages), all other asymmetric non-trigger vowel sets involve single occurrences of some combination.

| Asymmetric non-triggers | Harmonizing Feature | number of languages |
| :---: | :---: | :---: |
| e,o | [+ATR] | 26 |
| a | [+ATR] | 10 |
| I, ${ }^{\text {, }}$ | [-ATR] | 6 |
| e,o,a | [+ATR] | 5 |
| ${ }_{1}^{1}, \mathrm{U}_{1}$ | [+raised] | 3 |
| $\dot{\text { i }}$ | [+high] | 1 |
| e,o, $\varepsilon, \nu^{\text {c }}$ | [ $\pm$ ATR] | 1 |
| $\bigcirc, \mathrm{I}$ | [-ATR] | 1 |
| $\varepsilon$ | [-ATR] | 1 |
| $æ, \mathrm{y}, \mathrm{u}$ | [ $\pm$ back] | 1 |
| æ | [ $\pm$ back] | 1 |
| Total |  | 56 |

In the present approach, this could be described by restricting the class of head constraint on the heads of $[ \pm \mathrm{ATR}]$ spans to the following: a constraint against low vowels heading [+ATR] spans, against mid vowels heading [+ATR] spans and against high vowels heading [-ATR] spans, cf. (55). It should be kept in mind that all of these are related to more general articulatory markedness considerations (cf. also Archangeli \& Pulleyblank 1994). In an approach based on constraints on underlying representations, this instead leads to the familiar Duplication Problem and unexplained conspiracies.
(55) Markedness constraints on Span Heads in ATR harmony
a. *SpanHEAd $(+$ low $)(+$ ATR $)$
b. *SpanHEAd(-low,-high)(+ATR)
c. *SpanHead (+high)(-ATR)

### 4.6 Conclusion

In this chapter, I have shown that asymmetric non-triggers in vowel harmony are not a valid argument in favor of constraints on underlying representations. A parallel OT account can derive the data if the correct representations are provided. I argue that it is crucial for the representations to include a way of distinguishing a unique trigger of vowel harmony from other segments that the feature spreads to. In Headed Spans Theory this segment is the designated head of a feature span. Constraints on span heads can do double-duty. They ban segments from underived environments because they would have to head a simple feature span. Additionally, they ban certain segments from triggering
harmony, since in such a form they would need to head a complex feature span.
I have also argued that there are alternative proposals that enhance OT in other ways to derive such patterns. I have noted root strata and complex faithfulness constraints are empirically adequate in that they succeed in deriving asymmetric non-trigger patterns. Optimality Theory has thus several options to account for such data. On the other hand, I have shown that these approaches do not uphold the Richness of the Base and Output Drivenness of Optimality Theory, respectively.

Crosslinguistically, we have seen that asymmetric non-triggers show a strong tendency towards more articulatory marked vowels. The sample was composed mostly of AtlanticCongo languages from Africa for independent reasons. The preponderance of marked non-triggers could potentially be derived by restricting the set of possible constraints on span heads to certain articulatorily marked feature combinations.

## Chapter 5

## Richness of the Base Problems in Tonal Phonology

Tonal phonology is an area that is usually left uncovered when discussing Richness of the Base and/or Morpheme Structure Conditions. A recent dissertation (Gorman 2013) mentions vowel and consonant phonotactics but no phonotactics of tone or tonotactics. Similarly, the handbook article by Booij (2011) does not mention Morpheme Structure Constraints on tone at all.

This is in stark contrast to the actual practice in theoretical phonology. At least since the advent of autosegmental phonology, constraints on underlying representations have played a crucial role in the analysis of tone languages. From the beginning, underlying inventories of tonal sequences were not only posited but also restricted by additional constraints. The prime example is the obligatory contour principle (OCP), cf. Myers (1997) for an overview. Even theories explicitly couched in an Optimality Theory framework, such as Optimal Tone Mapping (Zoll 2003) still adopt tonal inventories.

This chapter deals with two kinds of challenges to the Richness of the Base principle in Optimality Theory. In the first part, I will provide a definition of a Richness of the Base problem that only rests on a set of constraints and input-output mappings. I provide a case study from Chungli Ao bound roots. This pattern has been argued to pose a challenge to Richness of the Base by Temsunungsang (2009) and I will show that it conforms to my definition of a Richness of the Base problem. However, I will also show that this problem vanishes if more complex representations are assumed, namely Hierarchical Morphoprosodic structure. Positional constraints on tone, relativized to prosodic domains, play a major role in this analysis.

The second part of this chapter deals with a different challenge to Richness of the Base: the established theory of Autosegmental Phonology that includes language-specific constraints
on underlying representations. I will demonstrate that these are not needed in a parallel OT framework with positional markedness constraints. Richness of the Base can be maintained and a reanalysis is possible.

Both parts are unified by the solution they provide. The distribution of tones is governed by positional constraints on tones in certain prosodic domains. No reference to underlying tonal melodies is needed. At least for the analyses presented here, an even stronger claim is possible. No constraint solely refers to the tonal tier and restricts tonal melodies independent of their association status. ${ }^{1}$

The remainder of this chapter is organized as follows. I will start out by formalizing a definition of the Richness of the Base problem, based on a generalized version of the Escape Lemma (Prince 1998) in section 5.1. I will continue by presenting some complex tonal data from Chungli Ao, which pose a severe challenge to Richness of the Base and provide a solution based on Morphoprosodic Hierarchical Structure and positional markedness constraints in section 5.2. Section 5.3 includes a discussion and a reanalysis of data that has been used as an argument for melodic and positional tonal inventories. I also argue that any universal restriction on underlying tonal melodies proposed in the framework of Autosegmental Phonology has been faced with counterexamples from very early on. Before concluding in section 5.5 , I will show some crosslinguistic data on general and category specific constraint on tone in monomorphemic domains in section 5.4.

### 5.1 Richness of the Base Problems

In order to discuss Richness of the Base problems in tonal phonology, we have to establish what constitutes such a problem. We have already seen specific Richness of the Base problems in previous chapters. Researchers have felt the need to posit restrictions on the input even in an Optimality Theory framework. I will use the definition given in (1). In this section, I will justify this definition, starting from a simpler problem that I term the Missing Identity Mapping problem.
(1) Richness of the Base problem

A potential input /A/ would map to an output [B] under the constraint ranking that has been established based on other input-output pairings and
a. empirically /A/ maps to [C] and
b. $[B]$ is independently well-formed and
c. the mapping $/ \mathrm{A} /[\mathrm{C}]$ is properly less faithful than the mapping $/ \mathrm{A} /[\mathrm{B}]$.

[^58]The definition of a Richness of the Base problem in (1) references three entities: an input $/ A /$ and two outputs $[B]$ and $[C]$. The first two subclauses refer to the attestedness of the output [B]. The problem can be substantially simplified if we assume identity of /A/ and [C]. The mapping of an input to an identical output as a property of grammars has been termed the Identity Map Property. A grammar has this property if and only if an input that is also a well-formed output, maps to an identical output.
(2) Identity Map Property (cf. e.g. Prince 1998; Tesar 2013:4)

A well-formed output [A], when used as an input /A/, maps to itself.

Magri (2018) provides a somewhat different formal characterization of this property as the reflexivity axiom. It states that the presence of an non-identical mapping yielding output [B] implies an identical mapping of the input /B/ to the identical output [B]. This means that an identity mapping is only enforced for an output $[B]$ if the output is independently attested for some other input.
(3) Reflexivity Axiom (Magri 2018)

If the candidate set contains a candidate that maps $/ A /$ to $[B]$ with $\Re(A, B)$, it also contains the identity candidate $/ \mathrm{B} /[\mathrm{B}]$ with $\Im(\mathrm{B}, \mathrm{B})$, where $\Im(\mathrm{B}, \mathrm{B})$ is the identity correspondence relation among the segments of $/ B /$ and $[B]$.

It might seem that this property is trivial and there should be no empirical challenges to it. Some segmental patterns have already been established in the literature to pose a challenge to a grammar with the identity map property, e.g. vowel chain shift and phonological derived environment effects (cf. Tesar 2013). These do not automatically pose a challenge to Richness of the Base or require restrictions on the input. Consider however the following example from Koryak Labials. All data are from Abramovitz (2019, 2021).

In Koryak, /v/ and /w/ are neutralized to /w/ in syllable final position but contrast otherwise. Compare the data in (4). [w] can occur syllable-initially as well as syllablefinally, as in the form [wanaw]. [v] on the other hand can only occur syllable-initially, as in the form [vutqəvut] but never syllable-finally.
(4)

| Distribution of /v/ and /w/ in Koryak |  |  |
| :---: | :---: | :---: |
|  | non-syllable-final | syllable-final |
| [w] | wutku | wajaw |
|  | 'here' | 'word' |
|  | wanavatək | jewjew |
|  | 'to speak' | 'partridge' |
|  | ewejul?etke |  |
|  | 'not scared' |  |
|  | tekuwinnetyi |  |
|  | 'I am helping you' |  |
|  | mətwapaquele |  |
|  | 'we two searched for fly agaric' |  |
| [v] | vutqəvut | - |
|  | 'darkness' |  |
|  | vatqolRen |  |
|  | 'last' |  |
|  | yeveq |  |
|  | 'if' |  |
|  | uvik |  |
|  | 'body' |  |
|  | mətətvamək |  |
|  | 'we two were' |  |

This can be easily captured by an OT grammar that bans syllable-final /v/. In all other contexts, neutralization is blocked by a lower ranked identity constraints Ident(son). A tableau for a potential input with a syllable-final $/ \mathrm{v} /$ is given in (5). The faithful candidate (5-a) cannot become optimal due to high ranked markedness constraint $\left.{ }^{*}\right]_{\sigma}$ against syllable-final $[\mathrm{v}]$. Therefore, the optimal candidate ( $5-\mathrm{b}$ ) repairs this violation by changing /v/ into [w].
(5) Evaluation of $/ \mathrm{w} /-/ \mathrm{v} /$ neutralization

| I: /wajav/ | $\left.*_{\mathrm{v}}\right]_{\sigma}$ | IDENT(son) | $*_{\mathrm{v}}$ |
| :---: | :---: | :---: | :---: |
| a. [wajav] | *! |  | * |
| ¢ ${ }_{\text {b }}$ [ [wajaw] |  | * |  |

As expected under this ranking, we find alternation from a syllable-finally /w/ to a syllable-initial /v/. If the context changes such that the markedness constraint does not apply anymore, /v/ can surface. Data are given in (6). The isolated forms [wanaw] and [jewjew] in ( 6 -a,b) end in a [w]. The derived forms [wanavatək] and [jewjevu] on the other
hand show a $[\mathrm{v}$ ] in the same position.
(6) /w/-/v/ alternation in Koryak
a. [wanaw] vs. [wanavatək] 'word' vs. 'to speak'
b. [jewjew] vs. [jewjevu]
'partridge' vs. 'partridges'
c. [jewjew] vs. *[jewjewu]

Another set of data is missing however. There is no non-alternating syllable-final /w/. This is exemplified by the data in (6-c). There is no pair of forms where the both the underived form [jewjew] and the derived form *[jewjewu] show a [w] in a resyllabifying position. Under the currently assumed ranking, we would expect an input /w/ to map to $/ \mathrm{w} /$ both if in the coda and if resyllabified as an onset. Onset /w/ cannot be banned since onset / w/ is a well-formed output. Richness of the Base seems to run into a problem here. Allowing input / w/ predicts a non-alternating pattern that is absent from the empirical data. I refer to this problem as the missing identity mapping problem. A definition is given in (7).
(7) Missing Identity Mapping Problem

An input / $\mathrm{A} /$ that would be an independently well-formed output $[\mathrm{A}]$ does not map to itself. Instead, it maps to a different output [B].

Even though the specific formulation of this problem has to my knowledge no precedence in the literature, the space of possible solutions in conservative Optimality Theory has already been provided as the Escape Lemma in Prince (1998). Prince notes that an identity map from / $\mathrm{A} /$ to $[\mathrm{A}]$ will not occur any violations for faithfulness constraints, since it is by definition the most faithful candidate. Therefore, only the high ranking of a markedness constraint banning $*[A]$ in the output can achieve a non-identical ranking. It has to be ranked above all relevant faithfulness and markedness constraints, as stated in the definition of the Escape Lemma in (8).
(8) Escape Lemma (Prince 1998)

Because the only disadvantages of the Identity map are in the Markedness constraints, in order to prefer the mapping $/ \mathrm{A} /[\mathrm{B}]$ to $/ \mathrm{A} /[\mathrm{A}]$, some constraint *A from the set of markedness constraints must dominate
a. all of faithfulness constraint that militate against the change from /A/ to $[B]$ and
b. all markedness constraints that prefer $[\mathrm{A}]$ over $[\mathrm{B}]$.

In order to put the Escape Lemma to work in Koryak labials, we have to figure out which markedness constraint to use. Consider the output in (9-a) and (9-b). The outputs are segmentally identical in the relevant aspects. Both include a syllable-initial glide /w/ preceding a vowel $/ \mathrm{u} /$. The only difference is the presence of a morpheme boundary in (9). Since this is the only reliable difference, the markedness constraint in (10) has to indirectly make reference to it. Following Hierarchical Morphoprosodic domain, a morpheme domain is defined by being directly dominated by the same prosodic word, cf. also chapter 1 and chapter 3 for more details on direct domination.
a. *jewjew-u
'partridges'
b. wuktku
'here'
$\left.*_{\mathrm{w}}\right]_{\omega, \mathrm{DD}}$
Count one-violation for every [w] that is at the right edge of a domain directly dominated by the same prosodic word.

The correct generalization for the distribution of Koryak labials is thus the following: There is a contrast everywhere but in syllable-final and morpheme-final position. In morpheme-final position we find $/ \mathrm{v} /$ unless it is syllable-final, where we find $/ \mathrm{w} /$. In syllable-final position we always find /w/. This can be easily translated into our constraint ranking by introducing the new constraint between the markedness constraint and the faithfulness constraint. ${ }^{2}$ In (11), $\left.{ }^{*}\right]_{\omega, \mathrm{DD}}$ excludes the faithful candidate (11-a), with $\left.{ }^{*}\right]_{\sigma}$ becoming inactive since the segment in question is now in the onset. Candidate (11-b) repairs /w/ to $[\mathrm{v}]$ and becomes optimal.

[^59]| Escape by $\left.*_{\mathrm{w}}\right]_{\omega, \mathrm{DD}}$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| I: /jewjew/-/u/ | $\left.*_{\mathrm{v}}\right]_{\sigma}$ | $\left.*_{\mathrm{w}}\right]_{\omega, \mathrm{DD}}$ | IdEnt(son) | ${ }^{\mathrm{v}}$ |
| a. [jew.je.w] ${ }_{\omega} \mathrm{u}$ |  | *! |  |  |
| b. [jew.je.v] $]_{\omega} \mathrm{u}$ |  |  | * | * |

So far, the discussion was focused on mapping inputs to identical outputs. The problem however can be extended to other cases, where a less faithful mapping needs to be preferred, even though a competing more faithful candidate would also result in a well-formed output. Two concepts need to be defined in order to achieve such a definition: relative faithfulness and independent well-formedness. Relative faithfulness must be based on a subset relation. ${ }^{3}$ Crucially, simply counting faithfulness violations does not suffice, since a higher number of faithfulness violations in lower ranked faithfulness constraints does not outweigh a single violation of a high ranked faithfulness. If an output is less marked and properly less faithful than another output for some given input, then it is harmonically bound for this constraint set and can never become optimal. The definition is given in (12).

Relative Faithfulness
A mapping /A/ $[B]$ is properly less faithful than a mapping /A/ [C] iff the faithfulness violations of $/ A /[B]$ are a proper subset of the faithfulness violations of $/ \mathrm{A} /[\mathrm{C}]$.

Additionally, we have to establish what a well-formed output is. Note that this definition cannot simply refer to the occurrence of some structure in the language, since it might derive from the input under discussion. Therefore, I will instead define the concept of independent well-formedness in (13).

Independent well-formedness of Outputs
An output $[\mathrm{C}]$ is independently well-formed for a mapping $/ \mathrm{A} /[\mathrm{C}]$ iff it is the grammatical output for some other input $/ \mathrm{B} /$ that is not identical to $/ \mathrm{A} /$.

With these terms established, we are able to formulate a more general version of the Escape Lemma. Instead of comparing an identity mapping to a non-identity mapping, we now compare a properly more faithful mapping / $\mathrm{A} /[\mathrm{C}]$ to a properly less faithful mapping $/ A /[B]$. In order to favor the less faithful mapping /A/[B] to become optimal, a high ranked markedness constraint against the output [C] has to be introduced. Otherwise, the less faithful candidate $[B]$ cannot become optimal for the input $/ A /$. The definition is

[^60]provided in (14).

Generalized Escape Lemma
In order to prefer a mapping $/ A /[B]$ to a mapping $/ A /[C]$, where $/ A /[B]$ is properly less faithful than a competing mapping / $\mathrm{A} /[\mathrm{C}]$ and $[\mathrm{C}]$ is an independently well-formed output, a markedness constraint $*[\mathrm{C}]$ has to dominate
a. all of faithfulness constraint that militate against the change from $/ \mathrm{A} /$ to $[\mathrm{B}]$ and
b. all markedness constraints that prefer [C] over [B].

From the Generalized Escape Lemma, we can work back to a general characterization of a Richness of the Base problem. Whereas the Generalized Escape Lemma describes a solution to generate a correct ranking, the term Richness of the Base problem refers exactly to the problematic situation that we started out with. A properly less faithful mapping is attested and preferred over another more faithful mapping, even though the output of the less faithful mapping is independently attested, cf. the definition in (15) repeated from (1).
(15) Richness of the Base problem

A potential input / $A /$ would map to an output $[B]$ under the constraint ranking that has been established based on other input-output pairings and
a. empirically /A/ maps to $[\mathrm{C}]$ and
b. $[\mathrm{B}]$ is independently well-formed and
c. the mapping /A/[C] is properly less faithful than the mapping /A/[B].

### 5.2 Abstract Tonal Inventories in Chungli Ao

Chungli Ao (Sino-Tibetan; India) features a set of intricate tonal alternations in bound verbal roots. As Temsunungsang (2009) points out, these alternations pose a problem to an Optimality Theory framework with Richness of the Base. In this section, I will demonstrate the Richness of the Base problem and provide an analysis based on recursive prosodic words that builds on the analysis developed in Temsunungsang (2009) but avoids restrictions on possible input melodies as well as cyclic or stratal computation. Instead, recursive prosodic words allow indirect reference to morphological affiliation in the phonological output. This allows us to differentiate candidates that would otherwise be identical or similar. The analysis does not include any constraint that solely refers to the tonal melody.

The Richness of the Base problem in bound roots can be introduced by looking into a
certain tonal class of verbs and the alternation between the isolated past tense form and the imperative form. Note that bound roots are augmented by a prefixed /a/- to conform to word minimality restrictions. The imperative suffix $-/ \overline{\mathrm{a}} \mathrm{y} /$ on the other hand is part of the class of trigger suffixes (Bruhn 2009) and induces alternations on the tone of the verb. Take for example the verb root /cum/ 'escape'. As shown in (16), the tone of the verb root is high (H) in isolated past form but low in the imperative form before the suffix $-/ \bar{a} \eta /$.

> /H/-/L/ alternations in Chungli Ao
a. ácún
escape[PST]
'escaped'
b. cùn-āg
escape-IMP
'escape!'

A very sensible assumption would be to posit a / $\mathrm{HL} /$ underlying tone for this root, since it alternates between a high tone and a low tone. This is problematic, however, since another class of verbs has exactly this pattern in the isolated past form. Take for example the verb root $/ \mathrm{ku} /$ 'to hit'. In its isolated form, the verb root bears a low tone and the epenthetic initial vowel /a/ bears a high tone.

$$
\begin{align*}
& \text { /ákù/ }  \tag{17}\\
& \text { hit[PST] } \\
& \text { 'hit' }
\end{align*}
$$

As far as tones are concerned, this is an instance of the missing identical mapping problem. The mapping /HL/ [HL] needs to be excluded, even though [HL] is a well-formed output form and the mapping does not induce any relevant faithfulness violations. Evidently, in this case we need to find an additional possible input tone that can account for the tonal alternations without neutralizing with another pattern. In subsection 5.2.4, I will show that $/ \mathrm{LH} /$ is one such an input representation.

Another example of a Richness of the Base problem appears when we look at the [L.M] imperative pattern again. Such an [L.M] pattern is unattested in isolated past forms of bound roots. This means that even though $[\mathrm{L} . \mathrm{M}]$ is attested in the imperative forms, a mapping /L.M/ to [L.M] must be excluded in the isolated past forms. This is an instance of the missing identity mapping problem because the identity map /L.M/[L.M] needs to be excluded, even though [L.M] is an independently well-formed output.

### 5.2.1 Tonal Alternations in Bound Roots

This section will give an overview of the alternations in bound verb roots in Chungli Ao. Note that we only focus on bound roots. As explained above, bound roots occur with an epenthetic initial /a/- in the past tense form that is otherwise unmarked. Before the set of trigger suffixes ${ }^{4}$, root tones undergo alternations. An overview of the alternations before trigger suffixes is given in (18). They occur before imperative, progressive and perfect suffixes. ${ }^{5}$

Bound roots before trigger suffixes

| class | root | isolated [PST] | IMP | PROG | PRF |
| :---: | :---: | :---: | :---: | :---: | :---: |
| mid | cum | ācūm | cūmāy | cūmtáùr | cúmúkū |
|  | 'drink' | M.M | M.M | M.H.L | H.H.M |
| weak low | juk | ájùk | jūkāy | jūktáùr | júkúkū |
|  | 'sell' | H.L | M.M | M.H.L | H.H.M |
| strong low | mun | ámùn | mùnnāy | mùntáùur | mùnúkū |
|  | 'sit' | H.L | L.M | L.H.L | L.H.M |
| weak high | kuk | ákúk | kūkāg | kūktáùr | kúkúkū |
|  | 'win' | H.H | M.M | M.H.L | H.H.M |
| strong high | kuy | ákúg | kùgāp | kùntáùr | kùgúkū |
|  | 'dry' | H.H | L.M | L.H.L | L.H.M |

In each column of the table in (18), there is only a very restricted set of tonal patterns. In the isolated past forms, only a mid-mid pattern /M.M/, a high-low pattern /H.L/ and a high-high pattern $/ \mathrm{H} . \mathrm{H} /$ are allowed. In each of the other columns only two patterns are allowed. In the imperative and the progressive forms, the root can either be mid toned /M.M/ or low toned /L.M/. The perfect forms coincide here: whenever the imperative/progressive tone is /M.M/, the root is high toned in the perfect forms. Whenever the imperative/progressive tone is /L.M/, the root in the perfect form is also low toned. Mid tones /M.M/ isolated past forms always take the mid respective high tone on roots in derived forms. The mapping from low or high toned isolated roots to root tones in derived forms is not as simple. In fact, we are faced with a many-to-many relation. A low toned root in an isolated past form can correspond either to a mid toned root or to a low toned root in the imperative form. The same is true for high toned isolated roots and/or progressive and perfect forms. This is visualized in the diagram in (19). The derived patterns $\mathrm{M} / \mathrm{M} / \mathrm{H}$ and $\mathrm{L} / \mathrm{L} / \mathrm{L}$ can be mapped to either of the two isolated root melodies /L/ or / $\mathrm{H} / \mathrm{and}$ vice versa.

[^61](19) Mapping of isolated root tones to tones before trigger suffixes


As Temsunungsang (2009) notes, there is a syllabic property unifying the weak low class and the weak high class. The root syllable is always checked. Chungli Ao distinguishes tonally between checked syllables, i.e. those that end in a plosive, and smooth syllables, i.e. those that do not. Checked syllable cannot bear a mid tone. However, as should be clear from the table in (18), this restriction only holds in underived environments. Note for example how the checked root syllable /kúk/ bears a high tone in isolation [ákúk] but a mid tone before the progressive suffix [kūktáùrr]. This is a first hint at the fact that tonal restrictions on isolated forms are different from the ones in derived forms, an idea that will be taken up in the discussion of checked syllables in subsection 5.2.4.

A different set of suffixes is available in Chungli Ao, which I will refer to as non-trigger suffixes. These suffixes do not cause the tonal alternations found with trigger suffixes. Non-trigger suffixes do not block the insertion of epenthetic initial /a/-. Additionally, bound roots mostly show up with the same tone that they show in isolation. The only tonal process affecting root tones here is a raising process that applies to underlyingly low toned roots and assigns a mid tone if they are followed by a high tone on the suffix. The non-trigger suffixes, on the other hand, change their tones depending on the tonal class of the root. When the future marker -/tsu/ follows a mid tone verb it surfaces with a high tone, otherwise with a low tone. The pattern for the nominalizer -/pa?/ is slightly different. It surfaces with a low tone only after weak and strong high tone verbs but as a high tone otherwise. Note that the non-trigger suffixes undergo tonal processes depending on the tone of the isolated form of the root. This is in stark contrast to the trigger suffixes that cross-classify the tonal root classes and tonal processes are applied accordingly. An overview of the data is given in (20). These data show that the tone that we saw in isolated past forms plays a role in the realization of suffixal tones.
(20)

Bound roots before non-trigger suffixes

| class | root | isolated [PST] | FUT | NMLZ |
| :--- | :--- | :--- | :--- | :--- |
| mid | ca | ācā | ācātsú | ācāpá? |
|  | 'call' | M.M | M.M.H | M.M.H |
| weak low | juk | ájùk | ájùktsù | ájūkpá? |
|  | 'sell' | H.L | H.L.L | H.M.H |
| strong low | ku | ákù | ákùtsù | ákūpá? |
|  | 'hit' | H.L | H.L.L | H.M.H |
| weak high | cupp | ácúp | ácúptsù | ácúppà? |
|  | 'cry' | H.H | H.H.L | H.H.L |
| strong high | kuy | ákúy | ákúytsùi | ákúypà? |
|  | 'dry' | H.H | H.H.L | H.H.L |

The third set of suffixes that I will discuss here are overwriting suffixes. These suffixes neutralize all tonal distinctions in verb roots and force their tone onto the root. The suffix tone can be any tone, but the resulting root tone is either mid or high. The conative suffix -/tāy/ assigns a mid tone and is itself mid toned. The causative suffix /tsúri / and the reciprocal suffix -/túp/assign a high tone to the root and are themselves high toned. A large class of suffixes are low toned and assign a high tone to the root. This includes the abilitive suffix -/tùt/. Note that in these cases initial /a/ epenthesis does not take place. Crucially, the resulting tonal melodies are identical to the ones found in isolated past roots, as shown in (21). ${ }^{6}$
(21) Bound roots before overwriting suffixes

| class | root | isolated <br> [PST] | CON | CAUS <br> /RECP | ABIL/DES /ChANCE |
| :---: | :---: | :---: | :---: | :---: | :---: |
| mid | rug | ārū | rūgtāŋ | rúgtsúf | $\dagger$ |
|  | 'drink' | M.M | M.M | H.H | H.L |
| weak low | juk | ájùk | jūktāg | júktúp | záknə̀† |
|  | 'sell' | H.L | M.M | H.H | H.L |
| strong low | mun | ámùn | mūntāy | múntsúf? | múm\î̀ $\dagger$ |
|  | 'sit' | H.L | M.M | H.H | H.L |
| weak high | cup | ácúp | cūptāy | cúptúp | cúrptùt |
|  | 'win' | H.H | M.M | H.H | H.L |
| strong high | kug | ákúg | kūgtāy | $\dagger$ | $\dagger$ |
|  | 'dry' | H.H | M.M | H.H | H.L |

[^62]
### 5.2.2 Recursive Prosodic Words as Tonal Domains

Temsunungsang $(2009,2017)$ proposes to account for the difference between overwriting and non-trigger suffixes by assuming that they attach to different prosodic positions. In this subsection, I will describe the main idea and expand it to explain other tonal differences. The concrete analysis in this subsection will focus on non-trigger suffixes, since they show fewer interactions.
(22) Prosodic position of non-trigger and overwriting suffixes


We can analyze non-trigger suffixes as occupying a higher prosodic position. Therefore they cannot influence the tone of the root. Additionally, this allows for a lower domain inside these forms. In this domain, the same tonal sequences that are allowed in isolated roots can appear. In the same domain, /a/ epenthesis applies. I will assume that this domain is the prosodic root word, i.e. the minimal prosodic word that includes the root.

Overwriting suffixes attach at a lower prosodic position. Therefore, the patterns produced here are identical to the ones attested in isolated forms. Similarly, trigger suffixes will be included in the prosodic root word. Additionally, minimality requirements can be satisfied by these suffixes without the need for an epenthetic initial /a/-.

Non-trigger suffixes seem to not alter the base they attach to considerably. I conclude that they are not included into any prosodic structure present in the prosodic root word. This also implies that they remain unfooted. Crucially, I will assume that constraints on tones not included in any foot differ between the domains of non-minimal and prosodic root words. Unfooted syllables that are directly dominated by a non-minimal word can only be high or low toned. Mid tones do not occur here.

In implementation, this can be accomplished by three undominated markedness constraints. In order to prohibit any interaction, tones and feet must be banned from spanning the boundary between the prosodic root word and the higher suffixes. This is accomplished by the CrispEdge ( $=\mathrm{CE}$ ) constraints in (23-a) and (23-b).

Constraints on non-minimal prosodic words
a. CrispEdge(Tone,Prosodic Word) $(=\mathrm{CE}(\mathrm{T}, \omega))$

Count one violation for any tone that is associated to two different syllables that are each directly dominated by a different prosodic word.
b. CrispEdge(Foot,Prosodic Word) $(=\mathrm{CE}(\mathrm{Ft}, \omega))$

Count one violation for any foot that is that is directly dominated by two different prosodic words.
c. $\mathrm{M} \rightarrow \mathrm{Ft}_{\omega \mathrm{NMIN}, \mathrm{DD}}$

Count one violation for any mid tone that is associated to a syllable that is directly dominated by a non-minimal prosodic word and not included in a foot.

An example evaluation is given in (24), where the root /kù/ 'hit' combines with a potentially mid-toned non-trigger affix -/ts $\overline{\mathrm{u}} /$. The mid tone cannot surface, as candidate (24-a) fatally violates the constraint against unfooted mid tones in non-minimal prosodic words. ${ }^{7}$ Any imaginable repair strategy is excluded. The mid tone cannot be rescued by simply associating to the foot inside the prosodic root word, since this would mean that the mid tone would span the prosodic word boundary, which fatally violates the CrispEdge(Tone,Prosodic Word) constraint. Similarly, refooting to include the affix inside the root foot would violate the similar constraint CrispEdge(Foot,Prosodic Word). This is exemplified by candidate (24-d). Instead, a change in the tonal specification of the affix becomes optimal, as shown in candidate (24-b). ${ }^{8}$

[^63](24) Non-trigger affixes add their tones and do not interact

| I: | CE(T, $\omega$ ) | $\mathrm{CE}(\mathrm{Ft}, \omega)$ | $\mathrm{M} \rightarrow \mathrm{Ft}_{\omega \mathrm{NMIN}, \mathrm{DD}}$ |
| :---: | :---: | :---: | :---: |
|  |  |  | *! |
|  |  |  |  |
|  | *! |  |  |
| d. |  | *! |  |

Note that the above constraint on non-footed tones cannot hold in the whole prosodic word, as explained in the next subsection. Three further general markedness constraints are needed to derive the few changes that we find in in roots and affixes here. The first is a general constraint against a low tone flanked by two high tones. The constraint in (25-a) seems to hold for all verb words in Chungli Ao (Temsunungsang 2009:135). Similarly, the sequence M.L does not occur in Chungli Ao verbs. This can be explained as a part of the markedness constraint family against tonal contours with a non-maximal distance between the two tone targets. Similarly, an $\mathrm{OCP}(\mathrm{H})$ constraint is needed to lower the nominalizer -/par/ after fully high toned roots. ${ }^{9}$ All these violations will be repaired by changes in tone specifications.

[^64](25) General markedness constraints affect non-minimal prosodic words
a. *H.L.H

Count one violation for any low tone that occurs between two high tones.
b. *M.L

Count one violation for any sequence of a mid tone followed by a low tone.
c. $\quad \mathrm{OCP}(\mathrm{H})$

Count one violation for each pair of adjacent syllables that are associated to different high tones.

I have shown in this section that non-triggering affixes can be derived by assigning them to a higher prosodic position. They remain unfooted, resulting in a loss of tonal distinctions and a vulnerability to undergo tonal repair mechanisms. Due to the prosodic word boundary and CrispEdge constraints, they cannot influence the tones of roots.

### 5.2.3 (Non-)Footing in Prosodic Root Words

There are two main differences between minimal and non-minimal prosodic words in Chungli Ao. They refer to footing and the status of non-footed syllables. The first one is the tone of non-footed syllables. Whereas it cannot be a mid tone in non-minimal prosodic words, this is licit in prosodic root words. On the other hand, high tones cannot occur on unfooted syllables here. This can be formalized as the constraint in (26), which bans high tones from unfooted syllables. Secondly, prosodic specification of affixes plays a role in this evaluation. The constraint, CrispEdge $(\mathrm{H}, \mathrm{Ft})$ prohibiting spreading of a high tone over a foot edge will turn our crucial here.
$\mathrm{H} \rightarrow \mathrm{Ft}_{\sqrt{\omega}}$
Count one violation for every high tone that is associated to a syllable that is dominated by a prosodic root word and not associated to a foot.
CrispEdge( $\mathrm{H}, \mathrm{Ft}$ ) $(=\mathrm{CE}(\mathrm{H}, \mathrm{Ft}))$
Count one violation for every high tone that is associated to two syllables that are dominated by different feet.

This constraint explains why the progressive suffix -/táùr/ can only be preceded by a mid tone or a low tone. The suffix itself already comes with a foot from the lexicon. This foot will not be altered, since it is protected by faithfulness constraints on footing. ${ }^{10}$

Consider the tableau in (27). A potential input with a high toned root /kúk/ combining

[^65]with the progressive suffix -/táùrr/ is evaluated. Since the suffix comes with a foot from the lexicon, the root is now unfooted in a prosodic root word, as shown in (27-a). In this position a high tone fatally violates the constraint $\mathrm{H} \rightarrow \mathrm{Ft}_{\sqrt{\omega}}$ that requires H tones in prosodic root words to be linked to a foot. Simply linking a high tone to both the root and the suffix, as done in candidate (27-c), fatally violates the constraint against tones straddling foot boundaries. This problem cannot be repaired by changing the foot structure, since this would violate the faithfulness constraint on feet. This is shown in candidate $(27-\mathrm{d})$. Including the root in the foot satisfies the constraint on high tones but nevertheless cannot become optimal. Instead, candidate (27-b) is the winner in this tableau. This is accomplished by changing the tonal specification of the root, which only violates the low ranked faithfulness constraint on tone.

Trigger suffixes impose tonal restrictions by preventing footing

|  | $\mathrm{H} \rightarrow \mathrm{Ft}_{\sqrt{\omega}}$ | $\mathrm{CE}(\mathrm{H}, \mathrm{Ft})$ | Faith(Ft) |
| :---: | :---: | :---: | :---: |
|  | *! |  |  |
|  |  |  |  |
|  |  | *! |  |
| d. |  |  | *! |

This strategy provides an analogical solution for the mid toned imperative suffix $-/ \bar{a} \bar{y} /$. Since it also only allows for low tones or mid toned syllables preceding it, we can assume that it is underlyingly footed in a defective foot that only includes one syllable. This means that a high tone can be excluded by simply banning them from occurring on unfooted syllables.

The evaluation for a root combining with the imperative suffix is provided in (28) and analogous to the evaluation of the progressive suffix. The faithful candidate (28-a) is
excluded due to the high tone on an unfooted syllable in a prosodic root word, fatally violating the constraint $\mathrm{H} \rightarrow \mathrm{Ft}_{\sqrt{\omega}}$. Spreading the high tone to the affix as in candidate (28-c) also violates the constraint CrispEdge( $\mathrm{H}, \mathrm{Ft}$ ) again. Finally, changing the footing is not allowed in candidate ( $28-\mathrm{d}$ ) even though it would satisfy the high ranked constraint $\mathrm{H} \rightarrow \mathrm{Ft}_{\sqrt{\omega}}$. Eventually, candidate (28-b) becomes optimal, since the faithfulness constraint for tone is ranked low enough.

Defective Feet on affixes prevent footing of roots

|  | $\mathrm{H} \rightarrow \mathrm{Ft}_{\sqrt{\omega}}$ | $\mathrm{CE}(\mathrm{H}, \mathrm{Ft})$ | Faith(Ft) |
| :---: | :---: | :---: | :---: |
|  | *! |  |  |
|  |  |  |  |
|  |  | *! |  |
|  |  |  | *! |

The perfective suffix -/úk $\overline{\mathrm{u}} /$ requires some further elaborations of the analysis. For one, it is bisyllabic, but it does not form a foot that is otherwise found in isolated past roots. This problem can be solved by assuming that it includes a high tone that is footed and a mid tone that is unfooted. Secondly, it can be preceded by a high tone or a low tone but not by a mid tone. This is due to the general unavailability of M.H.M sequences in Chungli Ao (Temsunungsang 2009:135). This has to be ranked above the constraint against doubly linked high tones in prosodic root words.
(29) Constraints on trisyllabic verb forms
a. ${ }^{*}$ M.H.M

Count one violation for every tonal sequence of a mid tone preceding a high tone that precedes another mid tone.

These data also require a more detailed version of the tonal faithfulness constraint. Changing a mid or high tone to a low tone must be worse than spreading a high tone. Changing a high tone into a mid tone must be allowed, since we find such a neutralization before the progressive and imperative suffix. It must also be possible for a mid tone to become a high or low tone, since such changes are allowed for affixes in higher prosodic positions.

We have thus shown that affixes with an underlyingly specified foot can reduce tonal constrasts on roots. The reasoning is analogous to the non-trigger affixes. Neither nontrigger affixes nor roots before trigger affixes are included inside a foot. Therefore they would violate markedness constraints on tonal contrasts in unfooted syllables if they kept all tonal distinctions.

This leaves the overwriting suffixes as the last group to be discussed. Their analysis is surprisingly simple. Up to now, I implicitly assumed that feet in Chungli Ao can only have the same tones that also occur in isolated roots. Let us make this restriction more explicit. Four markedness constraints are needed to exclude all illicit tonal specification on feet, leaving only H.L, H.H, and M.M as possible combinations. Low tones have to be banned from the first syllable of a foot. High tones on the other hand are required to be linked to the first syllable in a foot. The situation for mid tones is more complex. They can only occur if they are linked to both the first and the last syllable in a foot.
(30) Constraints on tonal foot specifications
a. ${ }^{*} \mathrm{~L} \rightarrow\left(\mathrm{Ft}^{\sigma}\right.$

Count one violation for every low tone that is associated to the first syllable of a foot.
b. $\mathrm{H} \rightarrow\left({ }_{\mathrm{Ft}} \sigma\right.$

Count one violation for every high tone that is associated to a footed syllable and not associated to the first syllable of a foot.
c. $\mathrm{M} \rightarrow\left({ }_{\mathrm{Ft}} \sigma\right.$

Count one violation for every mid tone that is associated to a footed syllable and not associated to the first syllable of a foot.
d. $\mathrm{M} \rightarrow \sigma)_{\mathrm{Ft}}$

Count one violation for every mid tone that is associated to a footed syllable and not associated to the last syllable of a foot.

Overwriting suffixes come without any prosodic prespecifications from the lexicon. They integrate into the prosodic root word. Depending on their tonal specification, they will trigger different patterns on the root. Their behavior is thus not different from roots. The
same markedness restrictions apply. ${ }^{11}$

Overwriting suffixes are as unmarked as isolated roots

|  | $\mathrm{H} \rightarrow(\mathrm{Ft} \sigma$ | ${ }^{*} \mathrm{~L} \rightarrow{ }_{(\mathrm{Ft}} \sigma$ | $\mathrm{M} \rightarrow \sigma)_{\mathrm{Ft}}$ | $\mathrm{M} \rightarrow{ }_{\text {Ft }} \sigma$ |
| :---: | :---: | :---: | :---: | :---: |
|  | *! |  |  |  |
|  |  |  |  |  |
|  |  | *! |  |  |
|  |  |  | *! |  |

Finally, some general restrictions on tonal spreading are needed. Tones never spread to the right. Trigger and overwriting suffixes can spread their tone to the root syllable. Isolated roots can spread their tone to a preceding epenthetic vowel. This can be accomplished by requiring each tone span to be headed by its rightmost tone. ${ }^{12}$ The constraint is given in (32). Feet seem to show a different pattern. Feet on affixes cannot integrate any preceding root material. Potentially present feet on roots need to be able to include overwriting affixes and epenthetic initial $/ \mathrm{a} /$. Under the assumption that epenthetic segments are incorporated into feet without violating any faithfulness constraint, feet thus need to allow extension only to the right. Feet are thus left headed. ${ }^{13}$

[^66](32) Constraints on Headedness of tonal spans and feet
a. SpanHeadRight(Tone)

Count one violation for each tone span that is not headed by its rightmost segment.
b. SpanHeadLeft(Foot)

Count one violation for each foot that is not headed by its leftmost syllable.

### 5.2.4 Richness of the Base Revisited

In the preceding sections, I have shown that given the right representations, the correct output forms can be derived, even if a wide range of inputs are allowed. This provides a specific solution to one Richness of the Base problem mentioned earlier. Why are there no isolated [L.M] roots even though this pattern is possible in affixed forms? The answer is that the outputs are not identical after all. The imperative forms include a defective foot on their final syllable, whereas the roots do not. Since the constraints on footed and unfooted syllables differ, $\left[\mathrm{L} .\left({ }_{\mathrm{Ft}} \mathrm{M}\right)_{\mathrm{Ft}}\right]$ with a footed final syllable is a grammatical output, but $\left[\left({ }_{\mathrm{Ft}} \mathrm{L} . \mathrm{M}\right)_{\mathrm{Ft}}\right]$ with both syllables in the same foot is not.

Let us now turn to the other Richness of the Base problem. What is the proper representation of a root that alternates between a high and a low tone? The answer is surprisingly simple. Any smooth syllable with a high tone, except if it already is $/ \mathrm{HL} /$ or also includes a mid tone. In fact, there are only three underlying two tone melodies that lead to a stable low tone. As expected, a syllable that already has an /HL/ contour will show up as low toned in isolation. Similarly, a single our double low tone will also yield a strong low toned root. Any root that contains a mid tone on a smooth syllable will yield a stable mid tone pattern. This way the lexicon is partitioned to assign a correct output to any melody in smooth syllables. This partition is summarized in (33). ${ }^{14}$
(33) Tonal partition of lexicon for roots with smooth syllables

| 1. $\backslash 2$. | H | M | L | $\emptyset$ |
| :--- | :--- | :--- | :--- | :--- |
| H | strong high | mid | strong low | strong high |
|  | H $\sim L$ | M | L | H $\sim L$ |
| M | mid | mid | mid | mid |
|  | M | M | M | M |
| L | strong high | mid | strong low | strong low |
|  | H $\sim L$ | M | L | L |
|  |  |  |  |  |

For checked syllable roots, we need to start out with formalizing the intuition that mid tones

[^67]cannot occur in the final syllable of an isolated bound roots. If we introduce a constraint on the tone of final syllables in prosodic root words, we can keep this generalization. Note that we only refer to footed syllables for reasons that will become clear in a moment.
\[

$$
\begin{equation*}
\left.\left.* \bar{\sigma}_{\text {checked }}\right)_{\mathrm{Ft}}\right]_{\sqrt{\omega}} \tag{34}
\end{equation*}
$$

\]

Count one violation for any footed mid toned checked syllable that occurs at the end of a prosodic root word.

This constraint must be high ranked. It will derive the generalization that any root with a checked syllable that has a mid tone in its tonal melody will realize this mid tone only if it occurs before a trigger suffix. In isolation, it will either get a [H.H] pattern as a default or realize the other low or high tone of its melody.

There are two remaining open issues concerning underspecification and checked syllable roots without any mid tone. For the latter we would expect, a strong high or strong low pattern on checked syllables. The most straightforward solution is to refer to checked syllables before trigger suffixes and exclude low tones from this position. Since root syllables before trigger suffixes are always unfooted, we can simply refer to checked unfooted syllables inside a prosodic word. These syllables are not allowed to bear a mid tone. ${ }^{15}$
$\grave{\sigma}_{\text {checked }} \rightarrow \mathrm{Ft}_{\sqrt{\omega}}$
Count one violation for any low toned checked syllable that is not associated to a foot and included in a prosodic root word.

A more serious problem are completely underspecified root syllables. In these cases, we would expect a mid tone in isolation, due to the markedness of a high tone. However, this mid tone would not be stable in smooth syllables, since no high tone epenthesis is required before a trigger suffix. One possible strain of analysis would assume that epenthesizing a high tone onto a foot head is somehow preferable to inserting a mid tone there.

In the last section, I have shown that large parts of the paradigm of Chungli Ao verbs can be derived in a parallel OT framework without the need for any restrictions on the input. This comes at a certain price. Phonological markedness constraints have to make reference to richer representations. This includes prosodic structure such as feet and recursive prosodic words but also headed autosegmental spans. Under these assumptions, Richness of the Base can be maintained.

[^68]
### 5.3 CoURs in Early Autosegmental Phonology

In early autosegmental phonology (e.g. Leben 1973; Goldsmith 1976) sequential Morpheme Structure Constraints were used extensively. In addition to an inventory of tonal autosegments, i.e. high tones, low tones, mid tones, etc., an inventory of underlying tonal melodies was posited. These were positive sequential constraints on underlying tones. Sometimes additional constraints on accentual diacritics were introduced. In this section, I will propose that such analyses can be translated into an OT system with ranked positional markedness constraints on tone. ${ }^{16}$ Richness of the Base can be maintained and no restrictions on the input are needed. This is highly relevant because tonal melody inventories are still often explicitly or implicitly assumed in OT analyses of tone. If combined with Hierarchical Morphoprosodic Structure, this approach also explains the often astonishing parallelism between monomorphemic domains and complex domains (Zoll 2003:234). Positional constraints can be relativized to monomorphemic domains or polymorphemic domains without any substantial difference.

The remainder of this section will first discuss melodic inventories as assumed in early Autosegmental Theory and show that a positional markedness analysis of Mende tone is possible and derives a wider range of empirical data than simple melodic inventories. This will be followed by a discussion of accentual inventories including analyses of Mee and Somali, which show the simplicity of a positional markedness account compared to accentual inventories. The section will conclude with a discussion of allegedly universal conventions in early Autosegmental Phonology. I will show that these were never meant to be universal and therefore violate the Richness of the Base just like other language-specific constraints on underlying representations do.

### 5.3.1 Melodic Tonal Inventories in Mende

The classic treatment of Mende (Mande, Sierra Leone) in Leben (1973) will serve to illustrate melodic inventories in early Autosegmental Phonology. ${ }^{17}$ This is arguably the most classic application of a melodic inventory. Leben $(1973,1978)$ posits five underlying tonal melodies (L,H,HL,LH,LHL) and illustrates this with the following data.

[^69]| melody | $1 \sigma$ | $2 \sigma$ | $3 \sigma$ |
| :---: | :---: | :---: | :---: |
| H | kó | pélé | háwámá |
|  | 'war' | 'house' | 'waistline' |
| L | kpà | bèlè | kpàkàlì |
|  | 'dept' | 'trousers' | 'tripod chair' |
| HL | ${ }^{\text {mbû }}$ | kénà | félàmà |
|  | 'owl' | 'uncle' | 'junction' |
| LH | ${ }^{\text {mbǎ }}$ | fàndé | ${ }^{\text {n }}$ dàvúlá |
|  | 'rice' | 'cotton' | 'sling' |
| LHL | ${ }^{\text {mbầ }}$ | nàhâ | nìkílì |
|  | 'companion' | 'woman' | 'groundnut' |

The basic idea of this part of the analysis here and later is that the tonal melody is simply mapped from left to right onto the syllables from left-to-right. This implies that tonal melodies are unlinked underlyingly and not prelinked. This association leads to a tonal crowding effect at the right edge of the word. As noted by Zoll (2003), tonal crowding due to directional melody mapping has two effects on the surface. Multiply linked tones are restricted to the edge that mapping ends at (the right edge in this case) and contour tones are restricted to the same edge.

## (37) Mende Melody Mapping

a. Tonal crowding creates contour tones at the right edge

b. Tonal crowding creates multiply linked tones at the right edge


This melodic inventory analysis can be translated into a set of ranked constraints. The tonal crowding effects of the mapping directionality can be mirrored by referring to an edgemost prosodic constituent and requiring tones to be linked to it. The size of the edgemost prosodic constituent can be derived from the maximal number of tones in a melody.

For Mende specifically, this means that there is a constraint restricting high tones to the initial foot $\mathrm{H} \rightarrow \# \mathrm{Ft}$. This relates to the empirical generalization that a high tone can only
occur in a root if it is associated to the word-initial foot. ${ }^{18}$ This means that in a word with three syllables, high tones will always surface if they can link to the initial foot. Three further constraints are needed to derive the deletion of high tones in longer forms. Falling tones are restricted to the final syllable by a constraint HL $\rightarrow \sigma \#$. This corresponds to the empirical generalization that falling tones only occur on the final syllable of a root. The last two constraints do not directly correspond to surface true generalizations. As for the constraint $\mathrm{LH} \rightarrow \# \sigma$ that bans rising tones on non-initial syllables, we can observe that rising tones only occur on initial syllables in a trivial sense in (36): they only ever occur on monosyllabic forms. Leben (1973:84) however, mentions two additional patterns, namely LH.H and LH.L, exemplified by [gǒtà] 'gutter', [gǒt̀̀] 'rope', [tǒlo] 'kola nut', [tǐkà] 'coal', [mǎnà] 'banana' and [běsí] 'pig'. Note that these are completely well-formed outputs under the present markedness constraints but pose a challenge if all association is automatic by convention. The last constraint is a formulation of the OCP constraint on high tones, combining insight from Leben (1973), Myers (1997), and Zoll (2003). ${ }^{19}$ It does not correspond to any obvious empirical generalization due to the possibility of multiply associated tones.
(38) Some undominated constraints on high tones in Mende
a. $\mathrm{H} \rightarrow \mathrm{FFt}$

Count one violation for every high tone that is not associated to a syllable that is in the initial foot of a prosodic word.
b. $\mathrm{HL} \rightarrow \sigma \#$

Count one violation for every syllable that is not final in a prosodic word and that is associated to a sequence of a high tone followed by the low tone.
c. $\mathrm{LH} \rightarrow \# \sigma$

Count one violation for every syllable that is not initial in a prosodic word and that is associated to a sequence of a low tone followed by a high tone.
d. $\quad \mathrm{OCP}(\mathrm{H})$

Count one violation for a pair of adjacent syllables that are associated to two separate high tones.

The application of these constraints can best be illustrated by evaluating an overspecified input. In (39), the input bears a /HLHL/ melody. It is not possible to associate all tones to the three syllables that are present without creating any contour tones. If the first / $\mathrm{HL} /$ sequence is linked to the first syllable as in candidate (39-b), the resulting falling tone

[^70]is on the first syllable. This is a clear and fatal violation of the positional markedness constraint on falling tones $\mathrm{HL} \rightarrow \sigma \#$, which requires falling tones to occur on the final syllable. It is necessary to delete at least one tone. This cannot be simply the final $/ \mathrm{H} /$. Consider candidate (39-d). If the tones are mapped one-to-one, the second high tone ends up on the third syllable, which is outside of the first foot. This violates the positional markedness constraint $\mathrm{H} \rightarrow \# \mathrm{Ft}$ which restricts high tones to the first foot. Linking the high tone additionally to the second syllable as in candidate (39-c) is also illicit, since it creates a rising tone on a non-initial syllable, which is exactly what is prohibited by $\mathrm{LH} \rightarrow \# \sigma$. The first low tone can also not be deleted. Simple deletion and one-to-one mapping would result in candidate (39-e) which fatally violates OCP(H), since the first and the second high tone of the underlying melody are now linked to a pair of adjacent syllables. This cannot be repaired by deleting one of the high tones, as done in candidate (39-f), since this would mean deleting more tones than necessary. Ultimately, candidate (39-a) becomes optimal. Here, the first high tone is deleted and a [L.H.L] melody results. No contour tones are created and the only high tone appears inside the first bisyllabic foot.
(39) Deletion of excessive high tones in Mende

|  H L H L <br>      <br> I:  ni ki li | HL $\rightarrow \sigma$ \# | $\mathrm{H} \rightarrow$ \#Ft | LH $\rightarrow$ \# $\sigma$ | OCP(H) | $\operatorname{Max}(\mathrm{T})$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  | $*$ |
| b. | *! |  |  |  |  |
|  |  | *! |  |  | * |
|  |  |  | *! |  | * |
|  |  |  |  | *! | * |
|  |  |  |  |  | **! |

Additional constraints are needed to derive the whole set of data. Obviously, there has to be a constraint against four tones linked to one syllable. Such a syllable does not occur in Mende. Additionally, a dipping contour tone (HLH) has to be excluded, since it does not occur on a single syllable, even in monosyllabic forms. Furthermore, mid tones have to be banned, since they do not occur in Mende.
(40) Further markedness constraints on tone in Mende
a. ${ }^{*} \sigma \rightarrow \tau_{4}$

Count one violation for each syllable that is linked to four or more tones.
b. ${ }^{*} \sigma \rightarrow \mathrm{HLH}$

Count one violation for each syllable that is linked to a sequence of High-Low-High tones.
c. ${ }^{*} \mathrm{M}$

Count one violation for every syllable that is associated to a mid tone.

All in all, a translation of a melody inventory to a set of (positional) markedness constraints is possible. A melodic inventory is formulated positively, but the first step is to describe the inventory based on negative constraints. In most languages, only certain types of tonal autosegments are allowed. In Mende for example, we do not find mid tones (M). This means that a complete OT analysis of Mende tone has to include a markedness constraint against mid tones.

Inventories necessarily have a finite upper bound of tones X per melody, in the case of Mende specifically a tonal melody cannot have four or more tones. This affects several markedness constraints. For one, it affects the maximal number of tones per syllable. Thus in Mende, syllables associated to four or more tones are banned by the constraint ${ }^{*} \sigma \rightarrow \tau_{4}$. This can be generalized to a constraint against a TBU being linked to $\mathrm{X}+1$ tones, where X is the upper bound of tones per melody in a melodic inventory approach. Additionally, the maximal number of tones per melody results in a restriction of contrastive tones to a certain window at one edge of the word. The tonal crowding effect mentioned above bans multiply linked tones from a certain window at a word edge. The size of this window is dependent on the maximal number of tones per melody, ranging from $\mathrm{X}-1$ to X , depending on the exact properties of the inventory. In Mende, the maximal number of tones in a melody is $\mathrm{X}=3$ and the window is foot-sized, containing two ( $=\mathrm{X}-1$ ) syllables. The corresponding positional markedness constraint $\mathrm{H} \rightarrow \# \mathrm{Ft}$ only restricts high tones to the initial foot. This is indirectly related to the inventory gap for HLH melodies, since such a melody would require a high foot linking outside of the foot.

The position of the tonal window is related to the mapping direction. The window for singly-linked contrastive tones is located at the edge where mapping starts. Therefore, a positional markedness constraint $\mathrm{H} \rightarrow \# \mathrm{Ft}$ restricts high tones to the initial foot and not the final foot. As mentioned before, the mapping direction also influences the position of contour tones. This is accomplished by the constraint $\mathrm{HL} \rightarrow \sigma \#$, falling tones are restricted to the final syllable. Since the more general restriction on high tones targets the foot at the opposite left edge, falling tones are effectively restricted to the final syllable of two syllable words. All other falling tones would necessitate a high tone linked outside the
first foot. This is not true for rising tones, which are restricted the initial syllable can in principle occur in words of any length. As mentioned above, they occur at least in monoand bisyllabic words, the latter being handled as an exception in an account based on melodic inventories. The constraints on contour tones have another interesting effects. Since simple contours are restricted to opposite edges, the complex contour LHL can only occur in monosyllabic words. Any LHL contour on a single syllable outside monosyllabic words would either violate $\mathrm{HL} \rightarrow \sigma \#$ or $\mathrm{LH} \rightarrow \# \sigma$. It is impossible for a syllable to be both final and initial in polysyllabic words.

The constraints on contour tones also partially derive the inventory gap *HLH in Mende. In bisyllabic roots, such a contour would either require a falling tone on the initial syllable or a rising tone on the final syllable. Both are banned by the positional markedness constraints on contour tones $\mathrm{HL} \rightarrow \sigma \#$ and $\mathrm{LH} \rightarrow \# \sigma$. In order to prevent the complex contour HLH to show up on monosyllabic words, the additional constraint ${ }^{*} \sigma \rightarrow \mathrm{HLH}$ is needed.
(41) Translation of melodic inventory to markedness constraints
Melodic Inventory Positional Markedness Mende Analysis


In this section, I have shown that melodic inventories are not necessary for an analysis of tonal languages such as Mende. Instead, a set of positional markedness constraints has been used to derive the same data. Note that this also allows us to derive the data that have to be marked as exceptional by pre-linking in an analysis of Mende tone based on melodic inventories. Rising tones can occur on the initial syllable, contrary to what a tonal crowding effect due to directional mapping would predict. This provides a further argument against a melodic inventory approach.

### 5.3.2 Positional Tonal Inventories in Mee and Somali

Early analyses of tonal phonology in an autosegmental framework have made use of accents assigned to underlying tone bearing units, cf. e.g. Goldsmith (1976) and Hyman (1981). These accents only occur in a restricted position underlyingly, which is assigned by special morpheme structure rules, and affect the position of the first association of tonal melodies. One example noted by Pulleyblank (1986) is that only one accent might occur per morpheme in some languages. However, Pulleyblank (1986) already notes that this notion of accentual diacritics is dependent on the idea that tone association is governed by universal conventions. If tonal association, i.e. linking and spreading, are governed by language specific principles, then diacritics and the relevant morpheme structure rules are not needed. Pulleyblank (1986), however, expands the notion of prelinking to derive such systems. In his theory, this entails lexical restrictions on possible pre-linked syllables. I will show that the language specific OT ranking of positional markedness constraints is enough to derive positionally restricted tone systems. In the remainder of this subsection, I will provide such analyses for restrictions on tone in Mee and Somali and generalize these analyses to an abstract way of translating accentual diacritics into positional markedness constraints.

Mee (Trans-New-Guinea; Indonesia) has a very restricted tonal inventory (Hyman \& Kobepa 2013). Tonal contrast between a high and a low tone is only present on the second mora. The first mora is always high toned and all moras starting from the third mora are low toned.

| Tone patterns in Mee |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :---: | :---: |
|  | $\mu \mu$ | $\mu \cdot \mu$ | $\mu \mu \cdot \mu$ | $\mu \cdot \mu \mu$ | $\mu \mu \cdot \mu \cdot \mu$ |  |  |
| $\# H L \ldots$ | bóù | údò | gáàbò | múmàì | máàkàdò |  |  |
|  | 'wind' | 'heavy' | 'quiet' | 'to finish' | 'true' |  |  |
| \#HHL... | bóú | 'udó | gáátì | búúmàì | áákàtà |  |  |
|  | 'to skin' | 'testicle' | 'ten' | 'to swim' | 'to belch' |  |  |

This a problem for a theory that only allows for melodic inventories. Both options are melodically a sequence of H and L. Very early on in Autosegmental Phonology, accents were introduced along with restrictions on their position (Goldsmith 1976; Hyman 1981). Accent marks are diacritic marks on a tone-bearing unit that are lexically specific and allow for exceptional association outside of the association conventions. In Mee, such an accent mark could be placed on the first or second mora and a HL pattern would be associated to it.

I will argue here that accent marks are not needed in an OT grammar if positional
markedness constraints are allowed. Positional Markedness constraints are enough to account for the positional tone inventory of Mee. Three markedness constraints are needed for this analysis. One constraint requires any high tone to be linked to the first foot in a prosodic word and another constraint requires the first mora to be linked to a high tone. A third constraint requires a one-to-one mapping between tones and moras. ${ }^{20}$
(43) Constraints needed to account for Mee tone distribution
a. $\quad \# \mu \rightarrow \mathrm{H}$

Count one violation for every mora that is
(i) leftmost in a prosodic word and
(ii) not associated to a high tone
b. $\mathrm{H} \rightarrow \# \mathrm{Ft}$

Count one violation for every high tone that is associated to a mora that is not included in the initial foot in a prosodic word.
c. One-to-one

Count one violation for every mapping between tones and moras that is not one-to-one.

When combined, these three constraints partition the lexicon into two. Inputs with one high tone or fewer result in a low tone on the second mora, whereas inputs with two high tones are mapped to outputs with a high tone on the second mora. The position of variation is simply where a high tone is allowed by the undominated markedness constraints but not required. No accent mark is needed here.

The tableau in (44) shows that surplus high tones are deleted, due to the positional markedness constraint that requires high tones to be linked to the first foot in a word, which - for independent reasons - is bimoraic. Therefore, the faithful candidate (44-a), with high tones outside of the first foot, is excluded. Similarly, deleting too many high tones, as in candidate (44-c) violates the faithfulness constraint $\operatorname{MAx}(\mathrm{T})$ fatally. Therefore, only candidate (44-b) with a high tone on the second mora will become optimal.

[^71]High tone deletion in Mee

|  | $\# \mu \rightarrow \mathrm{H}$ | $\mathrm{H} \rightarrow$ \#Ft | $\operatorname{Max}(\mathrm{T})$ |
| :---: | :---: | :---: | :---: |
|  |  | *!* |  |
|  |  |  | ** |
|  |  |  | ***! |

If there are not enough high tones in the input, they can be freely inserted, since $\operatorname{DEp}(T)$ is also ranked below the markedness constraints. This is shown in the tableau in (45). The faithful candidate (45-a) does not bear any high tones. Consequentially, its first mora does not bear a high tone either and the candidate therefore violates the constraint $\# \mu \rightarrow \mathrm{H}$ that requires the first mora to bear a high tone. In parallel to high tone deletion in (44), high tone insertion is required here. This necessarily violates the faithfulness constraint $\operatorname{DEP}(T)$ at least once, even in the optimal candidate (45-b). Nevertheless, inserting two high tones will violate the $\operatorname{DEP}(\mathrm{T})$ twice and thus candidate (45-c) cannot become optimal.

High tone insertion in Mee

|  | $\# \mu \rightarrow \mathrm{H}$ | $\mathrm{H} \rightarrow$ \#Ft | Dep(T) |
| :---: | :---: | :---: | :---: |
|  | *! |  |  |
|  |  |  | * |
| c. |  |  | **! |

In Somali (Afro-Asiatic; Djibouti, Ethiopia, Kenya, Somalia) words bear exactly one high tone with very few but systematic exceptions (Hyman 1981). The position of this high tone is also restricted. It can only occur on the final or penultimate mora of a word. Compare the data in (46).

Tonal inventory of Somali

|  | $\mu \mu$ | $\mu \cdot \mu$ | $\mu \mu \cdot \mu$ | $\mu \cdot \mu \mu$ |
| :--- | :--- | :--- | :--- | :--- |
| $\ldots$ LH\# | nàíl | dàqsó | 3òògsó | hàwèén |
|  | 'lamb.F' | 'stop!' | hurry! | 'woman' |
| $\ldots$. HL\# | él̀ | ínàn | qàalìn | dàméèr |
|  | 'dog' | 'son, boy' | 'young camel.m' | 'he-donkey' |

The pattern is thus almost an exact mirror image of the Mee pattern. There are two crucial differences. First, the final mora can be either high or low. This contrasts with the initial mora in Mee, which has to be high toned. Second, contrastive tones are not restricted to the penultimate mora of a word, as they are to the second mora of the word in Mee.

Two generalizations are true for the Somali data and they can be easily translated into positional markedness constraints. First, the final foot has to be linked to exactly one high tone. The corresponding positional markedness constraint is $\mathrm{Ft} \# \rightarrow \mathrm{H}_{1} .{ }^{21}$ Second, all high tones have to be linked to the word-final foot: $\mathrm{H} \rightarrow \mathrm{Ft} \#$. Additionally, a constraint is needed that bans multiply linked high tones, since these would allow a high tones to span across a foot boundary or link to both moras of the final foot. Finally, some low ranked markedness constraint has to distinguish between the two grammatical patterns. The constraint definitions are given in (47).

Positional Markedness Constraints on Somali Tone
a. $\mathrm{Ft} \# \rightarrow \mathrm{H}_{1}$

Count one violation for every foot that is
(i) final in the prosodic word and
(ii) not associated to exactly one high tone.
b. $\mathrm{H} \rightarrow \mathrm{Ft} \#$

Count one violation for every high tone that is not associated to the final foot in a prosodic word.
c. $\mathrm{H} \rightarrow \mu_{2}$

Count one violation for every high tone that is linked to multiple moras.

[^72]As shown in (48), these positional markedness constraints allow the grammar to deal with inputs that contain more than one high tone. Any high tone that is outside the final bimoraic foot violates the high ranked constraint $\mathrm{H} \rightarrow \mathrm{Ft} \#$, as in candidate (48-a). Additionally, this candidate violates the constraint requiring exactly one high tone in the final foot $\mathrm{Ft} \# \rightarrow \mathrm{H}_{1}$. This cannot be solved by linking the same high tone to both moras of the final foot, as in candidate (48-d), since this violates the constraint $\mathrm{H} \rightarrow \mu_{2}$ against multiply linked high tones. The final decision between candidates (48-b) and (48-c) is done purely based on some low ranked markedness constraint $M$ which prefers (48-b). Such a constraint could for example require a high tone to be linked to the final syllable.

High tone deletion in Somali

|  | $\mathrm{Ft} \# \rightarrow \mathrm{H}_{1}$ | $\mathrm{H} \rightarrow \mathrm{Ft} \#$ | $\mathrm{H} \rightarrow \mu_{2}$ | $\operatorname{Max}(\mathrm{T})$ | M |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | *! | * |  |  |  |
|  |  |  |  | ** |  |
|  |  |  |  | ** | *! |
| d. |  |  | *! | * |  |

As evident from the number of constraints, the analysis based on positional markedness is very simple. The essential ingredients are a positional markedness constraint that restricts high tones to a certain window and another constraint that requires some high tone to be linked to this window or a position inside it. Compare an analysis based on accentual diacritics. These minimally need to stipulate the presence of the accent, the rules governing its position and the inventory of tonal melodies including which tones bear accentual diacritics. A positional markedness analysis thus makes less assumptions than an account based on tonal and accentual inventories.

A final note on morphological tone. In Mee, as well as in Somali, morphologically triggered changes in tone can only add high tones to the right. This goes well with the idea that
they would alter the number of input tones but not the ranking of constraints. In Mee specifically, adding high tones to a root that already includes exactly one high tone, results in a change from a low tone on the second mora to a high tone.

### 5.3.3 Universal Conventions in Autosegmental Phonology

One possible objection to the classification of Autosegmental Phonology as a framework that makes heavy use of constraints on underlying representations is that the so-called conventions are supposedly universal and not language-specific. This section will serve to show that even in early autosegmental phonology, these conventions were not thought of as crosslinguistically exceptionless. I will also provide data that pose severe challenges to the assumption that any of these conventions are universal. ${ }^{22}$

The most well-known conventions of autosegmental phonology are probably the association conventions. These basically include an algorithm to associate unassociated tones to tonebearing units. This is not a direct constraint on underlying representations. Nevertheless, it needs to be discussed in order to understand the importance of other conventions. A version is given in (49).
(49) Universal Association Conventions (Pulleyblank 1986), cf. also Clements \& Ford (1979)

Map a sequence of tones onto a sequence of tone-bearing units,
a. from left to right
b. in a one-to-one relation

One property of this system is that the directionality of mapping is universal. This view was challenged by several researchers including e.g. Newman (1986) for Hausa and Odden (1981, 1984) for Karanga Shona. Directionality has later been treated as parametric. (e.g. by Lieber 1987:31).

Consider the following data from Hausa (Afro-Asiatic; Burkina Faso, Benin, Cameroon, Chad, Niger, Nigeria) in (50). Contrastive tones occur at the right edge and multiply linked tones occur at the left edge. This is the mirror image of the data seen in Mende in section 5.3.1. Newman (1986) shows that this can be easily derived by assuming right-to-left mapping.

[^73](50) Tone Mapping in Hausa (Newman 1986:250)
a. búhúnhúnàa
'sacks'
b. bàbbàbbàkú
'be well roasted'
c. cìnìkáyyàa
'mutual trade'

A third possibility has been proposed by Odden $(1981,1984)$ for Karanga Shona (Atlantic Congo, Zimbabwe) and was formalized by Yip (1988), Edge-In Association. ${ }^{23}$ The edgemost tones are associated to the first and last TBU and then tones spread inwards. Using the diagnostics from Zoll (2003) discussed before, multiply linked tones are restricted to a position that can be neither characterized as the left nor the right edge, cf. (51). In sum, mapping direction is thus not universal, even in early works on autosegmental phonology.

Shona non-assertive stems as evidence for edge-in mapping (Odden 1984:258)


An implicit assumption here and in most works is that underlying tones are not associated yet and their association is fully predictable. ${ }^{24}$ However, Leben (1973) already suggests that rising tones on initial syllable in Mende pose a challenge to this assumption (Leben 1973:84). Later, two solutions were suggested: either pre-linking of some tones in the underlying representation or accentual diacritics. The former option would entail that the floating status of underlying tones is not universal but at least language-specific. Pulleyblank (1986) convincingly argues in favor of such a theory and against accentual diacritics.

Accentual diacritics pose another problem for universality. From their inception, they have been pictured as only existing at the lexical level for some languages (Goldsmith 1976; Hyman 1981). Their presence or absence would thus be a language specific restriction on underlying representations. Similarly, their position is not universally fixed or free.

[^74]The analysis of Somali tone in Hyman (1981) for instance explicitly introduces a set of rules operating on underlying representations and introducing accentual diacritics. This produces another set of language-specific constraints on underlying representations.

Arguably, the earliest restriction on underlying melodic inventories is the Obligatory Contour Principle introduced in Leben (1973) and named by Goldsmith (1976). It states that underlying melodies cannot include two identical but distinct underlying tones in a sequence. It was first envisioned to work as a universal constraint on underlying representations.
(52) Obligatory Contour Principle (Leben 1973), as cited in Goldsmith (1976:63)

At the melodic level of the grammar, any two adjacent tonemes must be distinct.

However, Goldsmith (1976) already argues that this principle should be abandoned. He cites data from Etung (Atlantic-Congo; Cameroon, Nigeria) where contour tones occur at the right edge of a bisyllabic word, as in Mende, but additionally, the first and the second tone can be identical in some words. In (53), the contrasting melodies classified as LH vs. LHH and HL vs. HLL are given for bisyllabic and trisyllabic words.

|  | $2 \sigma$ | $3 \sigma$ |
| :---: | :---: | :---: |
| LH | ǹsí | bisóyé |
|  | 'fish' | 'spoon' |
| LLH | ǹsí | òròbé |
|  | 'mud' | 'beam' |
| HL | égòm | bíràmə̀ |
|  | 'jaundice' | 'me' |
| HHL | éfô | ésébè |
|  | 'cloth' | 'sand' |

Goldsmith (1976) takes the Etung data as evidence against a universal OCP. He notes that an analysis based on accentual diacritics could preserve the OCP as a constraint on underlying representations. However, this would imply introducing constraints on the underlying position of accentual diacritics. For the present purposes, this means trading the universality of one constraint on underlying representations for the universality of another.

Yet a different challenges is posed by trisyllabic nouns in Hausa. As mentioned above, Newman (1986) analyses Hausa using right-to-left mapping of tones. Nevertheless, certain
underived trisyllabic nouns seem to require OCP-violating melodies. As shown in (54), there are both LHH and HHL patterns. This is different from the predictions of a pure OCP-satisfying right-to-left association. Multiply linked tones occur to the left and to the right. This means that a right-to-left mapping analysis has to assume at least an LHH to account for the two high tones at the right edge. Note that an accentual diacritic approach becomes less likely, since the diacritic is only needed and allowed for LHH pattern nouns. This means there must be an additional, more specific constraint on underlying representation. Accent position has to be restricted depending on the tonal melody that is mapped to a noun.

```
Hausa OCP-violating melodies in underived trisyllabic nouns
HHL dzémágè: kíbíjà:
    'bat' 'arrow'
LHH dzimíná: màrák'í:
    'ostrich' 'calf'
```

In sum, all restrictions on the underlying representation assumed in autosegmental theory have been known to be language-specific for a long time. This means that they would always violate Richness of the Base. Therefore, they cannot be maintained in an version of OT that adheres to the Richness of the Base principle. This is not a problem. As argued in section 5.3.1 and 5.3.2, directionality and properties of the melodic inventory can be translated into surface true, rankable, positional markedness constraints. Directionality can generally be related to restricting certain tones or tonal contours to edgemost prosodic constituents. The floating status of underlying tones can be emulated by markedness constraint dominating faithfulness constraints, leading to predictable association. The OCP can be maintained as a rankable markedness constraint referring only to associated tones in surface representations. No constraint needs to refer to underlying tonal melody. Crucially, the rankability of these markedness constraints naturally allows for exceptions to their application.

### 5.4 Crosslinguistic Data

DoCoMD only includes nine entries of constraints on monomorphemic domains that concern tone. This is mainly a bibliographical bias, since in the description of tone systems, derived and underived contexts are often not clearly distinguished, as already criticized by Zoll (2003). ${ }^{25}$ Note that this does not pose a problem for Hierarchical Morphoprosodic Structure, since positional constraints can apply to any level of prosodic constituency.

[^75]Nevertheless, the tonal entries in DoCoMD show considerable diversity. In this section, I will describe the general properties of tonal CoMDs found in the database before focusing on category-specific constraints.

### 5.4.1 Tone in DoCoMD

In this section, I will briefly summarize the findings on tonal CoMDs in the database used in this dissertation. The six languages with tonal CoMDs in DoCoMD are given in (55). The languages are genealogically and areally diverse. Each language is from a different top-level family and three out of six macro-areas are represented. This supports the idea that tonal CoMDs are neither restricted to a certain language family nor to a certain macro-area.
(55) Languages with tonal CoMDs in DoCoMD

| Language | Top-level family | Macro-Area |
| :--- | :--- | :--- |
| Chungli Ao | Sino-Tibetan | Eurasia |
| Dena'ina | Athabaskan-Eyak-Tlingit | North America |
| Ejagham | Atlantic-Congo | Africa |
| Kere | Nuclear Trans New Guinea | Papuanesia |
| Ngiti | Central Sudanic | Africa |
| Tommo So Dogon | Dogon | Africa |

The properties of the CoMDs are also diverse. They apply in all possible domains, i.e. in affixes (Ejagham), in roots (Chungli Ao, Dena'ina, Kere, Ngiti, Tommo So Dogon) or in all morphemes (Ejagham). As far as their type is concerned, all possibilities are attested. Sequential constraints exist in Chungli Ao. Inventory constraints on tone are attested in Ejagham and Dena'ina. Positional constraints on tone are found in Ngiti and Tommo So Dogon and minimality/maximality constraints also exist in Tommo So Dogon. The table in (56) summarizes the properties of tonal CoMDs found in the database. Existing gaps are very likely due to the extremely small sample size. ${ }^{26}$

| Properties of tonal CoMDs in DoCoMD |  |  |  |
| :--- | ---: | ---: | ---: |
|  | root | affix | morpheme |
| Sequence | 1 | 0 | 0 |
| Inventory | 1 | 1 | 1 |
| Positional | 1 | 0 | 0 |
| Minimality/Maximality | 1 | 0 | 0 |

[^76]Several patterns are relevant for the discussion of Early Autosegmental Phonology. Constraints that affect affixes cannot receive a satisfying explanation if melodic inventories only exist for roots. In Ejagham, affixes can only ever bear a high, low and falling-to-low tone. This is different from roots, which can host more complex melodies. Nevertheless, all morphemes share the generalization that, if monosyllabic, they cannot bear a downstepped tone or a falling-to-mid tone (Waters 1989:54). In early Autosegmental Phonology this would entail two separate melodic inventories that only incidentally share some properties, unless further generalizations over melodic inventories are coded as constraints on underlying representations.

Similarly, constraints on the minimal or maximal occurrence of tones cannot be directly stated as positive sequential constraints. Instead, further constraints on tonal inventories are needed to constrain the melodic inventory if one aims at providing a generalization for this pattern. McPherson (2013) describes such a pattern for Tommo So Dogon, where all roots have to bear exactly one high tone stretch (cf. also Hyman 2006, 2009). In early Autosegmental Phonology, this would mean that melodies have to be constrained by excluding melodies such as L and HLH. The generalization receives a more straightforward explanation if Hierarchical Morphoprosodic Structure is assumed and minimality/maximality constraint can be directly relativized to root domains.

Another straightforward argument comes from Kere. In this language, tonal melodies can include high tones as their initial tone and monosyllabic words frequently show up with a high tone. However, isolated polysyllabic roots do not occur with a high tone on their first syllable (Rarrick 2017:99). This cannot be easily stated as a constraint on melodic inventories. A simple $/ \mathrm{H} /$ melody is needed for monosyllabic roots. Such a melody cannot be simply associated to polysyllabic roots. If an accentual diacritic is introduced, this would have to apply only to high tones. This positional generalization is of course easily describable in term of positional constraints. High tones need to link to the first foot in a root which is more important than a positional constraint against initial high tones. Note additionally that toneless initial root syllables can receive a high tone if it spreads from a preceding word. This can be modeled in a parallel OT framework by ranking a constraint on phrasal tone realization above the constraints mentioned before.

Finally, the pattern found in Dena'ina deserves special mention because it is unexpected from a markedness perspective. Dena'ina has a tonal inventory with low tones, mid tones and high tones; the latter, however, never occurs in simple roots. It only surfaces in negative verb forms (Lovick 2020:75). High tones are usually known to occur in prominent positions (de Lacy 2007). Roots are a prominent position, affixes are not (Beckman 1998). Banning high tones from a root therefore goes against expectation of positional faithfulness/markedness approaches. A brief discussion of the Dena'ina data is included in
chapter 2.

### 5.4.2 Category-specific Tonal Inventories

Category-specific effects on phonology have been known for some time but have been most thoroughly studied by Smith (2002, 2011). Smith also mentions lexical categories like the noun-verb distinction influencing tonal inventories. In this section, I will present data from Gã and provide a analysis of category-specific effects on tonal inventories. The analysis is based on Prosodic Category Inheritance, the idea that prosodic constituents can inherit categorial information from roots that they include.

In Gã (Atlantic-Congo, Ghana, Togo), all tonal contrasts occur in affixes and in noun roots, but tonal contrast is more restricted in verb roots. In verbs, bisyllabic roots cannot have a high tone on the first mora, followed by a high tone on the second mora. This surface pattern is allowed in nouns (Paster 2000). Some examples are provided in (57).

|  | verbs | nouns |
| :---: | :---: | :---: |
| H. ${ }^{\downarrow} \mathrm{H}$ | è-tfál ${ }^{\text {a }}$ | gó ${ }^{\text {wá }}$ |
|  | 'he mended' | 'guava' |
|  | è-bó ${ }^{\text {té }}$ e | jît ${ }^{\text {d }}$ ¢ ${ }^{\text {d }}$ |
|  | 'he entered' | 'head' |
| H.L | - | fótè |
|  |  | 'termite' |
|  |  | zénglè |
|  |  | 'roof' |

The basic idea of Prosodic Category Inheritance is that prosodic constituents can be specified for categorical information. This category is inherited from the morphological root included in the prosodic constituent. If a verb root is included in a prosodic word, it will be a prosodic word associated with the categorical feature verb. If a noun root is included in a prosodic word, it will be a prosodic word associated with the categorical feature noun.
(58) Prosodic Category Inheritance
a. The category of a prosodic constituent is inherited from a root that is included in the prosodic constituent.
b. Morphology $\longrightarrow$ Phonology


Coming back to Gã, the actual data are not simply referring to the tone of roots. Paster (2000) explicitly notes that [HL] structure are repaired to $\left[\mathrm{H}^{\downarrow} \mathrm{H}\right]$ in all prepausal contexts except in noun words. The exceptional [HL] pattern is not restricted to roots. ${ }^{27}$ Thus, the markedness constraint on all verbal prosodic words in (59) is enough to derive the data.
${ }^{*} \mathrm{HL}_{\omega \mathrm{V}}$
Count one violation for every sequence of a mora connected to a high tone followed by mora connected to a low tone that is included in a prosodic word with a categorial feature verb before a pause.

This constraint is violated only by prepausal [HL] in a prosodic word that bear a verb category feature. An example structure is given in (60-a). On the other hand, if such a tonal sequence occurs inside a prosodic word with a noun feature, the constraint is not violated. Such a structure is given in (60-b). This constraint thus reliably distinguished between the two structures. If it is ranked above the relevant faithfulness constraint, e.g. $\operatorname{DEP}(T)$, a repair strategy is triggered in verbs. Since the version of this constraint with a more general, category-unspecific domain is ranked below the faithfulness constraint, we do not expect any effects on nouns.

[^77](60) Violating and non-violating structure for ${ }^{*} \mathrm{HL}_{\omega \mathrm{V}}$

 bótè
$\omega_{\mathrm{N}}$
fótè

The analysis developed here can be extended to other cases of category-specific tonal inventories. In many languages, tone is predictable in verbs but contrastive in nouns. Smith (2011) discusses the cases of Tokyo Japanese (Japonic, Japan) and Mono (AtlanticCongo, Central African Republic, Democratic Republic of Congo). Marlo (2013) mentions additional 19 tone Bantu languages, where only verb tone is completely predictable and not contrastive. ${ }^{28}$ Korsah (2015) identifies and solves another Richness of the Base problem in Gã verbs: bisyllabic verbs cannot be underlyingly low-toned. Ngiti (Central Sudanic; Democratic Republic of Congo, Uganda) requires for verbs of the form VCVCV that all tones are identical (Kutsch Lojenga 1994:20). In Tommo So (Dogon; Mali) on the other hand, noun roots must have their high tone on a final TBU (McPherson 2013:77). As shown in section 5.2, Chungli Ao restricts the possible melodies for verb roots.

The pattern thus seems to be more typologically common than previously established. This provides evidence for reference to category features in phonology. As noted by Lohmann (2020b), category-specific effects on lexical tone have not been thoroughly examined. The sample presented in this section supplies a testing ground for the robustness of such patterns and their susceptibility to reanalysis as an effect of independent properties, such as lexical or syntactic distribution. Future research that denies the accessibility of category features in phonology needs to come up with possible schemata for reinterpretation of these data. In this section, I have established one possible representational solution to category-specific effects.

[^78]
### 5.5 Conclusion

In this chapter, I have shown that even though phonological practice for tonal analysis has continued to use restrictions on the input in the form of tonal inventories, such devices are not needed. Instead, positional markedness constraints are enough, given the right representations. In the analysis of Chungli Ao, we have seen that different positions in an output can be distinguished by their position in a foot and in non-minimal prosodic words or prosodic root words. Different positions can impose different markedness requirements.

As shown for Mende, Mee and Somali, positional markedness constraints can also derive the effects accomplished by tonal inventories and accentual diacritics in a straightforward manner. Any restriction on the input that has been argued to be universal by proponents of Autosegmental Phonology is seriously challenged by counterexamples. These restrictions thus do not escape the scope of Richness of the Base. Nevertheless, with the need for tonal inventories no longer present, the Richness of the Base principle can be upheld for tonal analyses in Optimality Theory.

## Chapter 6

## Discussion \& Conclusion

This chapter will provide a summary of the empirical and conceptual arguments that have been brought forward in this dissertation. In sum, Hierarchical Morphoprosodic Structure is specifically designed to account for effects traditionally linked to Morpheme Structure Constraints and can therefore provide an empirically adequate and explanatory account of these effects. Competing approaches face serious difficulties. The most severe problems are the diversity of attested domains, including both affixes and roots, and the typology of interaction between infixation and root domain constraints. The latter is particularly fatal, since all other theories overgenerate unattested patterns (Stratal OT, MSC) or undergenerate attested patterns (OT-OO). The arguments are summarized in the table in (1).

Properties of different theories

|  | HMS | Stratal OT | OT-OO | Comp. Faith. | MSC |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Richness of the Base | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | * |
| Output-Driven | $\checkmark$ | $\checkmark$ | ? | * | * |
| derived properties | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\times$ |
| diverse domains | $\checkmark$ | $\checkmark$ | * | * | $\checkmark$ |
| infixation interaction | $\checkmark$ | * | * | * | * |
| asymmetric non-trigger in VH | $\checkmark$ | $\nu$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |

It is possible that there is some combination of different competing approaches that accomplishes an empirical adequacy similar to Hierarchical Morphoprosodic Structure. However, it should be noted that any combinations of two of the alternative approaches discussed here, will not be able to maintain Output Drivenness. Additionally, I have not come across any combination of theories that easily derives the typology of interaction between infixation and root domain constraints.

In the following, I only discuss competing approaches that have been explicitly proposed to deal with this problems, even though other theories are discussed elsewhere in the thesis like the Root-Affix Faithfulness Metaconstraint in chapter 2, Harmonic Serialism in chapter 3, and classic Autosegmental Phonology in chapter 5. Classic Autosegmental Phonology faces the same problem that any approach based on MSCs encounters.

The remainder of this chapter discusses the arguments in more detail. In the next section, I will provide a summary of the arguments in favor of the theory proposed in this dissertation, i.e. Hierarchical Morphoprosodic Structure. The remaining sections, discuss the arguments against competing accounts of effects that have been traditionally explained by Morpheme Structure Constraints, in the following order: Stratal OT, OT-OO, Complex Faithfulness and constraints on underlying representations.

### 6.1 Hierarchical Morphoprosodic Structure

Hierarchical Morphoprosodic Structure is the theory presented in this dissertation. It provides a representational solution to the problems posed by effects traditionally accounted for by Morpheme Structure Constraints. Since Richness of the Base is adhered to, the Duplication problem does not arise. The Level Problem does not arise because all markedness constraints apply to the output. The Domain Problem is solved by recasting any morphosyntactic domain as a prosodic domain based on direct domination by prosodic words.

### 6.1.1 Richness of the Base and Output Drivenness

The main perspective for Hierarchical Morphoprosodic Structure is to explain all effects that have been attributed to Morpheme Structure Constraints as stemming from constraints on monomorphemic domains that apply to the output. This is achieved by relativizing markedness constraints to certain prosodic domains. These prosodic domains can be derived from prosodic constituents via the notions of domination and direct domination. The only further crucial representational addition is the headedness of autosegmental spans, following McCarthy (2004).

As far as Richness of the Base is concerned, Hierarchical Morphoprosodic Structure does not imply any violation. Relativized markedness constraints apply only to the output and do not restrict the set of possible inputs in any way. Similarly, since no new faithfulness constraints are introduced, Output Drivenness is maintained. How is that achieved? Output Drivenness is not only dependent on the linguistic data but also on the representational part of the analysis (Tesar 2013:14). By introducing more complex output representations, otherwise identical candidates can be differentiated, see the discussion of

Koryak in chapter 5 for more details.
Upholding Richness of the Base is conceptually desirable, since it provides a solution to the Duplication problem. In Hierarchical Morphoprosodic Structure, the same constraint can be relativized to different domains. Phonological generalizations that apply in several domains are not duplicated. All markedness constraints apply to the output.

### 6.1.2 Constraints on Derived Properties

Since relativized constraints in Hierarchical Morphoprosodic Structure apply to output representations, derived properties are available. This also means that the theory provides a solution to the Level Problem. Markedness constraints all apply at the same level.

Furthermore, Hierarchical Morphoprosodic Structure assumes an intricate relation between constraints on monomorphemic domains. This has important consequences. Following Downing (2006), this means that all prosodic size constraints can be reduced to conditions on prosodic constituents, here more specifically on prosodic words.

However, this advantage comes with a minor problem. In some cases, it seems that resyllabification counterbleeds constraints on monomorphemic domains, i.e. constraints on monomorphemic seem to make reference to syllable structure before resyllabification in complex forms. There are several possible solutions. Resyllabification could be blocked in these case, allowing for a simple analysis in terms of markedness constraints relativized to prosodic domains. This could be achieved by employing CrispEdge $(\sigma, \omega)$ constraints, which block resyllabification across prosodic word boundaries. A second possible solution is based on the idea that multidominance structures are available and interpreted in a certain way by the phonological constraints. A syllable could thus be part of two different prosodic words. Relativized constraints would then interpret this structures according to some convention, e.g. the first element belongs to the leftmost mother and the rest of the syllable belongs to the second mother, see also the example of Koryak in chapter 5.

### 6.1.3 Diversity of Domains

Hierarchical Morphoprosodic Structure allows reference to roots and affixes or both. This is accomplished by introducing the relation of direct dominance into the definition of constraint domains. In the prosodic structure in (2), all material directly dominated by a specific non-minimal prosodic word forms an affix domain. Similarly, material dominated by a prosodic root word, i.e. the minimal prosodic word that includes the root, forms a root domain, see also chapter 1 and chapter 3 for further discussion.
(2) Different kinds of prosodic words


Markedness constraints relativized to different monomorphemic domains thus provide a solution to the Domain Problem. This is a major advantage of the theory, which sets it apart from competing solutions inside Optimality Theory. Neither Stratal OT without a root stratum, nor OT-OO or Complex Faithfulness constraints achieve similar results.

### 6.1.4 Interaction with Infixation and Vowel Harmony

Hierarchical Morphoprosodic Structure can be easily restricted to derive the correct typology of the interaction between infixation and root domain constraints. Recall that roots always stay domains for root domain constraints, even if discontiguous through infixation. This means that a pattern where infixed roots are exempt from root domain constraints needs to be excluded by a theory of monomorphemic domains.

The crucial ingredient is a metaconstraint that requires a fixed ranking $M_{\pi, D D} \gg M_{\pi, D}$ between constraints on material directly dominated by prosodic constituents above constraints on material dominated by prosodic constituents. As shown in (3), infixes are representationally special. They are the only material that can be indirectly dominated by a prosodic root word, because they occur inside the root. Concretely, we must therefore rank a constraint on root material (directly dominated by $\sqrt{\omega}$ ) above the same constraint on root and infix material (indirectly dominated by $\sqrt{\omega}$ ). This excludes any pattern where a constraint applies to whole infixed form but not to the discontiguous root because this would require the opposite ranking. No other theory of constraints on monomorphemic domains can be easily restricted to generate only the attested interaction patterns, see chapter 3 for more details.
(3) Schematic representation of an infixed form with hierarchical morphoprosodic structure


The headed span part of the theory can be motivated by the asymmetric non-trigger pattern in vowel harmony. In this pattern, a certain vowel quality only occurs if derived by vowel harmony. It cannot occur in non-derived environments or as a trigger of vowel harmony. This is problematic for an OT analysis, where a trigger and a target of vowel harmony cannot be readily distinguished by markedness constraints. Headed spans provide a solution to this problem. Asymmetric non-triggers can only occur as non-heads, they are banned from heading a span by a designated markedness constraint. As a trigger of vowel harmony or in isolated contexts they would need to head their own feature span and are therefore repaired. A more detailed discussion is found in chapter 4.

### 6.2 Stratal OT

Stratal OT is a serial version of Optimality Theory that is characterized by a finite number of levels with different rankings. Each form has to undergo evaluation at each level in a fixed order. Any affix is assigned to a specific level.

I will discuss two different versions of this theory. The first version (cf. Trommer 2011) features a root stratum, i.e. roots and potentially affixes undergo a first round of evaluation before any morphological concatenation. This contrasts with the more standard version of Stratal OT, where the first step of evaluation applies after at least one affix has been potentially attached (Bermúdez-Otero 2012; Kiparsky 2015).

The former version of Stratal OT, where a root stratum is assumed, faces a conceptual problem. It allows for a language-specific ranking at the root level. In other words, there is a language-specific phonological sub-grammar that applies before any phonological and morphological operations to monomorphemic domains. This serves to derive effects traditionally attributed to MSCs. This makes Stratal OT with a root stratum conceptually indistinguishable from an account based on Morpheme Structure Constraints. Recall that exactly these were the three defining properties for MSC introduced in chapter 1, cf. also chapter 4 for a discussion. ${ }^{1}$ This problem does not apply in a version of Stratal OT without

[^79]a root stratum.
A further potential problem that Stratal OT with a root stratum shares with an MSC approach are constraints on derived phonological properties. As shown in chapter 2, constraints on monomorphemic domains can refer to derived phonological properties, such as syllable structure and contextual allophones. If some phonological properties are only derived at a later stratum, these properties cannot have an influence on the root stratum. This problem is less severe for a version of Stratal OT without a root stratum.

The version of Stratal OT without a root stratum faces a different problem however. If affixes only ever undergo phonological computation after being added to their base, they cannot form a proper domain for phonological generalizations to the exclusion of their base. This is especially problematic, since we have seen in chapter 2 that there are constraints on affix domains in different languages.

Independent from the inclusion of a root stratum, Stratal OT faces a major overgeneration problem when trying to derive the crosslinguistic typology of the interaction between root domain constraints and infixation. The main empirical finding of chapter 3 was that discontiguous roots always stay a domain for root domain constraints. Infixation never blocks the application of root domain constraints. There is no language, where root domain constraints apply to all roots, except when rendered discontiguous by infixation. This is a problem for Stratal OT, since such a language can be generated by a simple sequential and simultaneous interaction. If infixation happens either before or at the same stratum with root domain constraints, we expect discontiguous roots to be exempt from root domain constraints. ${ }^{2}$ This problem cannot be fixed by restricting reranking between different strata since it also arises from simultaneous intra-stratal interaction, see chapter 3 for a more detailed discussion.

### 6.3 Output-Output Faithfulness

Output-Output Faithfulness (Benua 1997) employs a non-derivational addition to Optimality Theory by extending the notion of correspondence to apply between the evaluated candidates and a designated output, the base. This base is part of the same paradigm as the candidates under evaluation. This correspondence relation is deployed by the so-called output-output faithfulness constraints. As shown in McCarthy (1998), these constraints can derive constraints on root domains in two different ways. The first possibility is that a change is carried out in the isolated root X and transferred to the derived form $\mathrm{X}-\mathrm{Y}$ where

[^80]the context for the change is not given anymore. A violation of a markedness constraint M is thus transparently repaired in an isolated root X, but this repair overapplies in a complex form X-Y where M is not violated anymore. Output-Output Faithfulness constraints are responsible for this preservation. The second possibility applies a change only to repair the violation of a markedness constraint M in the non-derived form C because the isolated root C is not protected by OO-faithfulness. However, if a complex form A-B violates the same constraint M , the repair is blocked by OO-faithfulness to a base A . This isolated root A does not violate the markedness constraint $M$ and can thus surface faithfully anyway. The OO-faithfulness constraint thus blocks a change from applying in derived forms. The two possibilities are schematized in (4).
(4) a. OT-OO preserving a change that applies in isolated roots

|  | $[\mathrm{X}-\mathrm{Y}]_{\boldsymbol{\nu} \mathrm{M}}$ |  | $[\mathrm{X}] * \mathrm{M}$ |
| :---: | :---: | :---: | :---: |
| IO-Faith | $\downarrow$ |  | $\downarrow$ |
|  | $[\mathrm{Z}-\mathrm{Y}]_{\boldsymbol{\nu M}}$ | $\leftarrow$ | $[\mathrm{Z}]_{\boldsymbol{\sim} \mathrm{M}}$ |

b. OT-OO blocking a change from applying in complex forms


OT-OO has the conceptual advantage that it upholds Richness of the Base and thus provides a solution to the Duplication problem. OO-faithfulness constraints prominently apply to the output only and do not make reference to the input in any way. This means that they do not restrict the set of possible inputs in a language-specific way. Their behavior with regard to Output Drivenness is more complex. They fall into the class of so-called non-stationary constraints (Tesar 2013:149) because the number of violations for a given constraint not only depends on the candidate under evaluation but also on the general constraint ranking. If we change the general constraint ranking, a different base output might be generated, which in turn will influence the number of violations assigned by an OO-faithfulness constraint. Tesar (2013:149) shows that another theory of non-stationary constraints, namely Sympathy Theory, derives non-output-driven patterns. It is thus very probable that OT-OO is also able to derive such patterns, since it was designed to deal with opaque phonological generalizations. Nevertheless, I will give OT-OO the benefit of doubt and not assume that it violates Output Drivenness. Instead, I will leave this question open.

As already noted by McCarthy (1998), one major advantage of OT-OO over MSCs, is its application to outputs. This solves the Level Problem. OT-OO can deal with constraints
on monomorphemic domains that make reference to derived phonological properties. Since OO-faithfulness constraints always apply to the output in a parallel OT system, all derived properties are present and available at their level of application. CoMDs referring to syllable structure or contextual allophones are thus completely unproblematic.

Regardless, OT-OO cannot derive the full crosslinguistic typology of constraints on monomorphemic domains. It faces a problem almost identical to the disadvantages of Stratal OT without a root stratum: the diversity of domains. OO-faithfulness constraints explicitly only refer to attested outputs as the designated base, i.e. only free forms can be domains for constraints on monomorphemic domains. This is in stark contrast to the result of the typological study in chapter 4. Affixes do in fact constitute domains for phonological generalizations. This problem cannot be easily overcome. A recent proposal by Rolle (2018) allows for reference to virtual bases, i.e. forms that are a potential but unattested output. It might be true that an extension of this approach, which would allow all monomorphemic forms as bases, is imaginable, but I am not sure that it is theoretically desirable, since it would undermine the predictive power of OT-OO and create additional conflicts for choosing a base or set of bases for any complex form. It is also not fully clear if such an approach would accomplish empirical adequacy.

A much more concrete instance of the Domain Problem arises when dealing with infixation. OT-OO cannot derive the Hebrew type of interaction between infixation and root domain constraints. Recall that in Hebrew, infixes do not become part of the root domain for purposes of root domain constraints. The inability of OT-OO to derive this pattern is related to its problems with affix domains. While alternations in the root can be blocked by referring to an isolated root as its base, affixes cannot be protected.

OT-OO can provide an analysis of the asymmetric non-trigger pattern in vowel harmony, yet not a very explanatory one. Recall that this refers to inventory constraints on certain vowel quality that only apply to isolated forms and the triggers of vowel harmony. The same vowel that is banned in these contexts might show up if derived by vowel harmony. I will only provide a rough sketch of an analysis here and point out a major shortcoming. The basic idea would be that an Ident-OO constraint on vowel quality is ranked high enough to prevent harmony from applying whenever a non-derived trigger underlyingly appears in a triggering position. If there is no such asymmetric non-trigger, no such blocking is needed. Consequently, the choice of the base needs to be determined by the presence or absence of a asymmetric non-trigger. This is the core of the explanatory problem. Recall that asymmetric non-triggers are characterized by their vowel quality, i.e. a purely phonological property. The base that each affix selects thus depends on a purely phonological property of said affix. Under an OT-OO analysis this is completely coincidental. The analysis therefore misses an important generalization.

### 6.4 Complex Faithfulness

Complex faithfulness constraint have been used by Mahanta (2008) to derive an asymmetric non-trigger pattern in Assamese vowel harmony. The basic idea is that standard faithfulness constraints can be combined with additional conditions on their target in the input and/or the output. This allows to indirectly restrict the input set in a language-specific way by mapping certain inputs irregularly to an unexpected output.

This means that Richness of the Base can be maintained because there need not to be constraints that restrict the input in the language specific constraint ranking. Output Drivenness, however, is not achieved. As argued in chapter 4, complex faithfulness constraints allows for the exclusion of more faithful rankings that are not motivated by markedness reduction.

When complex faithfulness constraints are employed in a parallel OT framework, they have the potential to derive constraints on phonologically derived properties. However, without additional representational devices they do not easily deal with a variety of monomorphemic domains. In Mahanta (2008) they are used to substitute for constraints on underlying representations. This means that they were not purposely designed to derive constraint on monomorphemic domains. As expected, they cannot easily refer to designated monomorphemic domains, e.g. the initial position of an affix.

Since Complex Faithfulness offers no solution to the Domain Problem, it is also incapable of deriving the typology of interactions between infixation and root domain constraints. Asymmetric non-triggers in vowel harmony, on the other hand, can receive an empirically adequate analysis, as shown in detail in chapter 4.

### 6.5 Constraints on Underlying Representations

Finally, constraints on underlying representations could be used to deal with constraints on monomorphemic domains, as proposed in early rule-based theories such as Halle (1959) and Chomsky \& Halle (1968). Obviously, such an approach does not solve the Duplication problem, since the same phonological generalizations often must be stated to apply over underlying representations and as an active part of the phonological grammar that governs alternations (cf. Kisseberth \& Kenstowicz 1977). Such an approach does not adhere to the Richness of the Base principle. Similarly, Output Drivenness cannot be maintained, since non-output-driven patterns such as chain shifts can be easily modeled as opaque rule interaction, e.g. counterfeeding in the case of chain shifts (Kirchner 1996). ${ }^{3}$

[^81]One major problem faced by an approach based on CoURs are phonologically derived properties. As shown in chapter 2, constraints on monomorphemic domains make reference to phonological derived properties. This is problematic for an account based on CoURs, since predictably derived phonological properties are not present in the underlying representations. A more detailed discussion can be found in chapter 2.

Constraints on underlying representations do not face any immediate problem with regards to the diversity of domains. As long as the underlying form of morphemes is marked for their root or affix status, both can serve as domains for phonological generalizations. When it comes to infixation however, approaches based on CoURs face a problem similar to the challenges of Stratal OT. If root domain constraints can be ordered both before and after infixation ${ }^{4}$, we expect a simple grammar, where infixation counterfeeds root domain constraints. This would produce a language where root domain constraints do not apply to discontiguous roots produced by infixation. Such a language is not attested. The approach cannot be easily restricted in order to get rid of the overgeneration problem. The Strong Domain Hypothesis (Kiparsky 1982, 1985) suggest itself. This would exclude the rule ordering needed to derive the unattested interaction pattern. Unfortunately, the Strong Domain Hypothesis has been faced with counterexamples from regular prefixation and suffixation almost from its inception. This would mean that a version of the Strong Domain Hypothesis relativized to infixation is needed. Such a move is highly stipulative and therefore undesirable, see chapter 3 for a more detailed discussion.

### 6.6 Conclusion

This dissertation has provided a treatment of effects traditionally attributed to Morpheme Structure Constraints. This treatment consisted of two parts. First, as presented in chapter 2 , a crosslinguistic study was conducted and yielded exploratory results and significant correlations. Second, a representational account of constraints on monomorphemic domains was proposed, additionally supported by the interaction of such constraints with infixation, vowel harmony, and tonal processes.

The crosslinguistic study on phonological constraints on monomorphemic domains, provided three conclusions. First, constraints on monomorphemic domains make reference to phonologically derived properties such as syllable structure and contextual allophones. This provides evidence against theories that try to explain all constraints on monomorphemic domains by constraints on underlying representations. Second, the study shows that phonological constraints can refer to affix domains and to root domains independently. Affix

[^82]and root domains pose a challenge to procedural theories without any precyclic phonological computation, such as OT-OO, Stratal OT without a root stratum and Cophonology theory as well as the Root-Affix Metaconstraint. Third, not all constraints are grounded in articulatory phonetics. Instead, some constraints can be explained by morphological parsing, as posited by Albright (2004). Morphological parsing is also a possible explanation for certain correlation between properties of constraints on monomorphemic domains. Vowel inventory constraints apply more frequently in non-roots in the controlled sample. Similarly, sequential constraints and positional constraints on consonants apply more often to roots. However, I also proposed alternative explanations based on typological bias or areal skewing.

Apart from the arguments based on the typological study, I have also examined the interaction of phonological constraints on monomorphemic domains with other morphological and phonological processes. In chapter 3, I have shown that the typology of interaction between infixation and root domain constraints can be accounted for by Hierarchical Morphoprosodic Structure in combination with a fixed ranking but poses a challenge to competing approaches. The fact that discontiguous roots always stay domains for root domain constraints, even if rendered so by infixation, is an empirical problem for procedural approaches. Such approaches predict a language where infixed roots are exempt from root domain constraints. Since no such language is attested, procedural approaches like Stratal OT with a root stratum, face an overgeneration problem.

The interaction of vowel inventory constraints with vowel harmony results in a pattern that I term asymmetric non-triggers, as shown in chapter 4. In this pattern, a certain vowel quality is banned from occurring in non-derived environments and as the trigger of vowel harmony. This vowel quality can only surface as an undergoer of vowel harmony. This pattern has been claimed to support constraints on underlying representations. Nevertheless, this pattern can be derived in a parallel OT system with Richness of the Base, given the correct representations, namely Headed Spans. The gist of the analysis is that certain vowel qualities can be banned from the head position of an autosegmental span. They can only occur as non-heads if derived by vowel harmony. Therefore, Hierarchical Morphoprosodic Structure can derive the empirical pattern just as well as other theories that have been proposed. This chapter also provides conceptual arguments against competing theories based on Richness of the Base and Output Drivenness.

The interaction with tonal processes has been analyzed in a detailed case study on Chungli Ao bound roots in chapter 5 . This complex pattern and its interaction with morphological structure has been argued to necessitate constraints on underlying tonal melodies. I have shown that this pattern can be analyzed using positional markedness constraints referring to tones and different prosodic domains, such as syllables, feet, and prosodic words. I
have also shown that melodic inventories from classic Autosegmental Phonology are not needed to explain the tonal patterns that have been used as classic arguments for such restrictions on underlying representations. The same kind of markedness constraints used for the analysis of Chungli Ao can be used to derive the data.

The results of this dissertation can also be seen as a confirmation of the hypothesis that all constraints on underlying representations can be derived by constraints on monomorphemic domains. Hierarchical Morphoprosodic Structure thus provides a solution to all three problems identified with Morpheme Structure Constraints by Kisseberth \& Kenstowicz (1977). The Duplication Problem is solved since Richness of the Base can be maintained. This means that one and the same phonological generalization can always be attributed to the same markedness constraint relativized to different prosodic domains. Similarly, the Level Problem is solved. All markedness constraints apply at the surface level. No further level of application is needed. Finally, the Domain Problem is solved. Markedness constraints can apply to a variety of prosodic domains. These domains, however, are all of the same kind, namely prosodic. This unites the analysis of constraints on monomorphemic domains with the analysis of other phenomena that make reference to metric and prosodic domains.

In chapter 1, I have also presented three hypothesis that are compatible with Hierarchical Morphoprosodic Structure: the Indirect Domain Hypothesis, the Domain Reference Hypothesis, and the No Static Phonotactics Hypothesis. In Hierarchical Morphoprosodic Structure, markedness constraints make reference to inclusion or the edges of certain domains of prosodic domains, which are only indirectly linked to morphosyntactic structure. There are no further static constraints on any domain that is not metric or prosodic. This contrasts with some other theories of the phonology-morphology interface. OT-OO allows faithfulness constraints that explicitly make reference to morphologically related outputs including to morphologically defined subword domains. Similarly, Stratal OT with a root stratum allows for static phonotactics applying to domains, directly defined by morphology.

Overall, one conclusion from the empirical part of this thesis is that language descriptions often omit information about phonological generalization that hold in non-derived environments or monomorphemic domains. Phonological descriptions often focus on the syllable and the word as the domains of phonological generalizations, sometimes adding phonological phrases. As I have shown in chapter 2, constraints on monomorphemic domains are a common pattern crosslinguistically. The underdescription problem is especially severe for tone. Generalizations on complex and simple forms are often conflated here.

Due to the broad scope of this dissertation, some questions have remained open. Future research might engage with the following prospects. One way to continue with the
crosslinguistic survey presented in chapter 2 is a duplication of the result with a larger typologically and areally controlled sample. This might decide the question if the correlation between different constraint properties in roots and non-roots are an effect of typological or areal bias. A diametrically opposed approach is to recheck all patterns on a one-to-one basis. This can involve two different methodologies. First, it might turn out that some patterns fall into a different category if more specific data are available. Furthermore, some patterns might be susceptible to reanalysis in terms of purely phonological non-prosodic domains. Even if probably not all data points will be affected by this, it might considerably expand or change the typology of constraints on monomorphemic domains. Second, a statistical approach is viable. As shown by Gorman (2013), some phonotactic generalizations that have been described for certain languages do not turn out to be statistically significant. This could be applied to more relevant patterns. Relatedly, future research could involve experimental methods to probe the cognitive reality of constraints on monomorphemic domains, e.g. by artificial grammar experiments (Zimmer 1969:cf. e.g.).

A further research perspective is that Hierarchical Morphoprosodic Structure could be applied to account for phenomena to other phenomena at the phonology-morphology interface. The difference between maximal prosodic words and non-maximal prosodic words could be used to define different affix domains and could explain phonological behavior of affix groups, e.g. the difference between cyclic and non-cyclic affixes. Similarly, the relation of direct domination could be employed to derive domains from smaller and larger metrical-prosodic constituents, such as syllables, feet, and phonological phrases. This could be used to provide a unified analysis for a variety of phonological problems, such as extrasyllabicity, extrametricality and the phonological properties of different clitics.

## Bibliography

Abramovitz, Rafael. 2019. Against the richness of the base: Two Koryak case studies. Handout presented at the IGRA Colloquium Series, Universität Leipzig.

Abramovitz, Rafael. 2021. Topics in the grammar of Koryak. Cambridge, MA: MIT dissertation.

Akinlabi, Akinbiyi. 2009. Neutral vowels in Lokaa harmony. Canadian Journal of Linguistics 12(2). 197-228.

Albright, Adam. 2000. The productivity of infixation in Lakhota. UCLA Manuscript.
Albright, Adam. 2004. The emergence of the marked: Root-domain markedness in Lakhota. LSA talk handout, Boston.

Ampofo, Joana Serwaa \& Ezer Rasin. 2021. True self-counterfeeding vowel harmony in akan serial verb constructions. In Proceedings of the 51st annual meeting of the north east linguistics society,

Andersen, Torben. 1986. Tone splitting and vowel quality evidence from Lugbara. Studies in African Linguistics 17(1). 56-68.

Andersen, Torben. 1989. The Päri vowel system with an internal reconstruction of its historical development. Journal of African languages and linguistics 11(1). 1-20.

Andersen, Torben. 1999. Consonant alternation and verbal morphology in Mayak (Northern Burun). Afrika und Übersee 82(1). 65-97.

Andersen, Torben. 2004. Jumjum phonology. Studies in African linguistics 33(2). 133-162.
Anderson, Stephen R. 1986. Disjunctive ordering in inflectional morphology. Natural Language $\xi^{2}$ Linguistic Theory 4(1). 1-31.

Archangeli, Diana \& Douglas Pulleyblank. 1994. Grounded phonology. Cambridge, MA: MIT Press.

Arsenault, Paul \& Alexei Kochetov. 2019. Two types of retroflex harmony in Kalasha: Implications for phonological typology. In Joan L. G. Baart, Henrik Liljegren \& Thomas E. Payne (eds.), Languages of northern Pakistan and surrounding regions: Linguistic studies dedicated to the memory of Carla Radloff, 40-76. Karachi: Oxford University Press.

Asher, Ronald E. \& TC Kumari. 1997. Malayalam. London: Routledge.
Bachra, Bernard. 2001. The phonological structure of the verbal roots in Arabic and Hebrew. Leiden: Brill.

Bakker, Dik. 2010. Language sampling. In Jae Jung Song (ed.), The Oxford handbook of linguistic typology, Oxford: Oxford University Press.

Beckman, Jill Noelle. 1998. Positional faithfulness. Amherst, MA: University of Massachuetts dissertation.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Amherst, MA: University of Massachusetts dissertation.
van den Berg, Helma. 1995. A grammar of Hunzib. München: Lincom Europa.
van den Berg, René. 1989. A grammar of the Muna language. Dodrecht: Foris.
Bermúdez-Otero, Ricardo \& Kersti Börjars. 2006. Markedness in phonology and in syntax: the problem of grounding. Lingua 116(5). 710-756.

Bermúdez-Otero, Ricardo. 2012. The architecture of grammar and the division of labour in exponence. In Jochen Trommer (ed.), The morphology and phonology of exponence, 8-83. Oxford: Oxford University Press.

Bickel, Balthasar. 2008. A refined sampling procedure for genealogical control. Language Typology and Universals 61(3). 221-233.

Blevins, Juliette. 2014. Infixation. In Rochelle Lieber \& Pavol Ştekauer (eds.), The Oxford handbook of derivational morphology, Oxford: Oxford University Press.

Blust, Robert. 2003. Thao dictionary. Taipei: Academic Sinica.
Booij, Geert. 2011. Morpheme structure constraints. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume \& Keren Rice (eds.), The Blackwell companion to phonology, 2049-2069. Oxford: Wiley Blackwell.

Breteler, Jeroen. 2017. Deriving bounded tone with layered feet in harmonic serialism: The case of Saghala. Glossa 2(1).

Bromley, Myron. 1961. The phonology of Lower Grand Valley Dani. 's Gravenhage: Martinus Nijhoff.

Bromley, Myron. 1981. A grammar of lower Grand Valley Dani. Canberra: Pacific linguistics.

Bruhn, Daniel. 2009. The tonal classification of Chungli Ao verbs. UC Berkeley PhonLab Annual Report 5(5). 64-80.

Bryant, Michael Grayson. 1999. Aspects of Tirmaga grammar. Arlington, TX: The University of Texas dissertation.

Carroll, Lucien Serapio. 2015. Ixpantepec Nieves Mixtec word prosody. San Diego, CA: University of California dissertation.

Carter, Allyson. 2000. Featural morphology: evidence from Muna irrealis affixation. Coyote papers: Working papers in linguistics from AZ 10. 1-17.

Casali, Roderic F. 2003. ATR value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. Linguistic Typology 7(3). 307-382.

Casali, Roderic F. 2012. [+ ATR] dominance in Akan. Journal of West African Languages 39(1). 33-59.

Cassimjee, Farida \& Charles Kisseberth. 1998. Optimal domains theory and Bantu tonology. In Theoretical aspects of Bantu tone, 33-132. Stanford, CA: CSLI.

Charney, Jean Ormsbee. 1993. A grammar of Comanche. Lincoln, NE: University of Nebraska Press.

Chomsky, Noam \& Morris Halle. 1968. The sound pattern of English. New York, NY: Harper \& Row.

Clements, George. 1984. Vowel harmony in Akan: A consideration of Stewart's word structure conditions. Studies in African Linguistics 15(3). 321-338.

Clements, George. 1985. Akan vowel harmony: a nonlinear analysis. In Didier L. Goyvaerts (ed.), African linguistics: Essays in memory of M.W.K. Semikenke, 55-98. Amsterdam: John Benjamins.

Clements, George \& Kevin Ford. 1979. Kikuyu tone shift and its synchronic consequences. Linguistic Inquiry 10(2). 179-210.

Cohen, Kevin. 2000. Aspects of the grammar of Kukú. München: Lincom Europa.

Cohn, Abigail C. 1992. The consequences of dissimilation in Sundanese. Phonology 9(2). 199-220.

Cole, Jennifer \& Charles Kisseberth. 1994. An optimal domains theory of harmony. Studies in the Linguistic Sciences 24(2). 101-114.

Creemers, Ava, Jan Don \& Paula Fenger. 2018. Some affixes are roots, others are heads. Natural Language $\xi^{3}$ Linguistic Theory 36(1). 45-84.

Crowley, Terry. 1998. An Erromangan (Sye) grammar. Honolulu: University of Hawai'i Press.

Crowley, Terry. 1999. Ura: A disappearing language of southern Vanuatu. Canberra: Pacific Linguistics.

Dench, Alan Charles. 1995. Martuthunira: A language of the Pilbara region of Western Australia. Canberra: Pacific Linguistics.

Downing, Laura J. 1998. On the prosodic misalignment of onsetless syllables. Natural Language $\mathcal{E}^{2}$ Linguistic Theory 16(1). 1-52.

Downing, Laura J. 2006. Canonical forms in prosodic morphology. Oxford: Oxford University Press.

Downing, Laura J \& Maxwell Kadenge. 2020. Re-placing PStem in the prosodic hierarchy. The Linguistic Review 37(3). 433-461.

Dryer, Matthew S. 2013. Prefixing vs. suffixing in inflectional morphology. In Matthew S. Dryer \& Martin Haspelmath (eds.), The world atlas of language structures online, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/26.

Dryer, Matthew S. \& Martin Haspelmath (eds.). 2013. WALS Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/.

Durie, Mark. 1985. A grammar of Acehnese. Dordrecht: Foris.
Elfner, Emily. 2016. Stress-epenthesis interactions in Harmonic Serialism. In John J McCarthy \& Joe Pater (eds.), Harmonic grammar and harmonic serialism, Sheffield: Equinox.

Embick, David. 2003. Locality, listedness, and morphological identity. Studia linguistica 57(3). 143-169.

Embick, David \& Alec Marantz. 2008. Architecture and blocking. Linguistic inquiry 39(1). 1-53.

Epps, Patience. 2008. A grammar of Hup. Berlin: Mouton de Gruyter.

Finley, Sara. 2008. Formal and cognitive restrictions on vowel harmony. Baltimore, MD: Johns Hopkins University dissertation.

Frajzyngier, Zygmunt. 1993. A grammar of Mupun. Berlin: Reimer Verlag.

Frisch, Stefan A, Janet B Pierrehumbert \& Michael B Broe. 2004. Similarity avoidance and the OCP. Natural language \& Linguistic theory 22(1). 179-228.

Gleim, Daniel, Gereon Müller, Mariia Privizentseva \& Sören E. Tebay. 2022. Reflexes of exponent movement in inflectional morphology. Natural language ${ }^{6}$ Linguistic theory Online First.

Glover, Bonnie Carol. 1988. The morphophonology of Muscat Arabic. Los Angeles, CA: University of California dissertation.

Goldsmith, John. 1976. Autosegmental phonology. Cambridge, MA: MIT dissertation.

Golston, Chris \& Richard Wiese. 1998. The structure of the German root. In Wolfgang Kehrein und Richard Wiese (ed.), Phonology and morphology of the Germanic languages, 165-186. Tübingen: Max Niemeyer Verlag.

González, Hebe Alicia. 2005. A grammar of Tapiete (Tupi-Guarani). Pittsburgh, PA: University of Pittsburgh dissertation.

Gorman, Kyle. 2013. Generative phonotactics. Philadelphia, PA: University of Pennsylvania dissertation.

Gouskova, Maria. 2003. Deriving economy: syncope in Optimality Theory. Amherst, MA: University of Massachusetts dissertation.

Green, Thomas Michael. 1999. A lexicographic study of Ulwa. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Halle, Morris. 1959. The sound pattern of Russian. The Hague: Mouton.

Hammarström, Harald, Sebastian Bank, Robert Forkel \& Martin Haspelmath. 2018. Glottolog 3.2. Jena: Max Planck Institute for the Science of Human History. Available online at http://glottolog.org, Accessed on 2021-04-29.

Hammarström, Harald \& Mark Donohue. 2014. Some principles on the use of macro-areas in typological comparison. Language Dynamics and Change 4(1). 167-187.

Hargus, Sharon \& Virginia Beavert. 2006. High-ranking affix faithfulness in Yakima Sahaptin. In Donald Baumer, David Montero \& Michael Scanlon (eds.), Proceedings of the 25th West Coast Conference on Formal Linguistics, 177-185. Somerville, MA: Cascadilla Proceedings Project.

Harley, Heidi. 2005. How do verbs get their names? denominal verbs, manner incorporation, and the ontology of verb roots in english. In Nomi Erteschik-Shir \& Tova Rapoport (eds.), The syntax of aspect: Deriving thematic and aspectual interpretation, Oxford: Oxford University Press.

Haspelmath, Martin. 2020. The morph as a minimal linguistic form. Morphology 30. 117-134.

Haurholm-Larsen, Steffen. 2016. A grammar of Garifuna. Bern: University of Bern dissertation.

Hayes, Bruce \& James White. 2015. Saltation and the p-map. Phonology 32(2). 267-302.
Hooper, Joan B. 1972. The syllable in phonological theory. Language 48(3). 525-540.
Hualde, José Ignacio. 1989. The strict cycle condition and noncyclic rules. Linguistic Inquiry 20(4). 675-680.

Hudu, Fusheini Angulu. 2010. Dagbani tongue-root harmony: A formal account with ultrasound investigation. Vancouver: University of British Columbia dissertation.

Huehnergard, John. 1997. A grammar of Akkadian. Atlanta, GA: Scholars Press.
van der Hulst, Harry. 2018. Asymmetries in vowel harmony: A representational account. Oxford: Oxford University Press.

Hyman, Larry M. 1972. A phonological study of $F^{\prime}$ 'fe'-Bamileke. Los Angeles, CA: University of California dissertation.

Hyman, Larry M. 1981. Tonal accent in Somali. Studies in African linguistics 12(2). 169-203.

Hyman, Larry M. 1993. Structure preservation and postlexical tonology in Dagbani. In Sharon Hargus \& Ellen M. Kaisse (eds.), Studies in lexical phonology, 235-254. San Diego, CA: Academic Press.

Hyman, Larry M. 2006. Word-prosodic typology. Phonology 23(2). 225-257.
Hyman, Larry M. 2009. How (not) to do phonological typology: the case of pitch-accent. Language Science 31(2). 213-238.

Hyman, Larry M. \& Niko Kobepa. 2013. On the analysis of tone in Mee (Ekari, Ekagi, Kapauku). Oceanic Linguistics 52(2). 307-317.

Iacoponi, Luca. 2018. Phonological agreement by headed feature correspondence: extending Correspondence Theory to output features nodes. New Brunswick, NJ: Rutgers University dissertation.

Inkelas, Sharon. 1990. Prosodic constituency in the lexicon. New York, NY: Garland.
Inkelas, Sharon. 1993. Nimboran position class morphology. Natural Language ${ }^{\mathcal{E}}$ Linguistic Theory 11(4). 559-624.

Inkelas, Sharon. 1995. The consequences of optimization for underspecification. In Jill N Beckman (ed.), Proceedings of the North East Linguistics Society, vol. 25, 287-302. Amherst, MA: GLSA.

Inkelas, Sharon. 1998. The theoretical status of morphologically conditioned phonology: a case study of dominance effects. In Geert Booij \& Jaap van Marle (eds.), Yearbook of morphology 1997, 121-155. Dordrecht: Springer.

Inkelas, Sharon \& Cemil Orhan Orgun. 1995. Level ordering and economy in the lexical phonology of Turkish. Language 71(4). 763-793.

Inkelas, Sharon \& Stephanie Shih. 2015. Tone melodies in the age of surface correspondence. In Proceedings from the 51st annual meeting of the Chicago Linguistic Society, vol. 51 1, 269-283. Chicago, IL: Chicago Linguistic Society.

Itô, Junko. 1988. Syllable theory in prosodic phonology. New York, NY: Garland.
Itô, Junko \& Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), The handbook of Japanese linguistics, 62-100. Oxford: Blackwell.

Itô, Junko \& Armin Mester. 2003. Weak layering and word binarity. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata \& Shin-ichi Tanaka (eds.), A new century of phonology and phonological theory, 26-65. Tokyo: Kaitakusha.

Itô, Junko \& Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In Toni Borowsky, Shigeto Kawahara, Takahito Shinya \& Mariko Sugahara (eds.), Prosody matters: Essays in honor of Elisabeth Selkirk, 280-303. London: Equinox.

Itô, Junko \& Armin Mester. 2021. Recursive prosody and the prosodic form of compounds. Languages 6(2).

Jaker, Alessandro \& Paul Kiparsky. 2020. Level ordering and opacity in Tetsótíné: a stratal OT account. Phonology 37(4). 617-655.

Janssen, Dirk, Balthasar Bickel \& Fernando Zúñiga. 2006. Randomization tests in language typology. Linguistic Typology 10(3). 419-440.

Jardine, Adam. 2016. Locality and non-linear representations in tonal phonology. Newark, DE: University of Delaware dissertation.

Jones, Patrick. 2014. Tonal interaction in Kinande: Cyclicity, opacity, and morphosyntactic structure. Cambrdige, MA: Massachusetts Institute of Technology dissertation.

Kaisse, Ellen M. 1993. Rule reordering and rule generalization in lexical phonology: A reconsideration. In Sharon Hargus \& Ellen M. Kaisse (eds.), Studies in lexical phonology, 343-363. San Diego, CA: Academic Press.

Kalin, Laura. 2021. Evidence from infix allomorphy on the fine timing of the morphosyntaxphonology interface. Handout from a talk presented for a colloquium at Universität Leipzig.

Kalin, Laura \& Nicholas Rolle. 2022. Deconstructing subcategorization: Conditions on insertion versus position. Linguistic Inquiry Advance publication.

Kaplan, Aaron. 2011. Harmonic improvement without candidate chains in Chamorro. Linguistic Inquiry 42(4). 631-650.

Kaplan, Aaron. 2018. Positional licensing, asymmetric trade-offs and gradient constraints in harmonic grammar. Phonology 35(2). 247-286.

Kaplan, Aaron. 2019. Overshoot in licensing-driven harmony. Phonology 36(4). 605-626.
Kiparsky, Paul. 1982. Lexical morphology and phonology. In The linguistic society of Korea (ed.), Linguistics in the morning calm: Selected papers from SICOL-1981, 3-91. Seoul: Hansin.

Kiparsky, Paul. 1985. Some consequences of lexical phonology. Phonology yearbook 2(1). 85-138.

Kiparsky, Paul. 2015. Stratal OT: A synopsis and FAQs. In Yuchau E. Hsiao \& Lian-Hee Wee (eds.), Capturing phonological shades within and across languages, 1-45. Newcastle: Cambridge Scholars Publishing.

Kirchner, Robert. 1996. Synchronic chain shifts in optimality theory. Linguistic Inquiry 27(2). 341-350.

Kisseberth, Charles \& Michael Kenstowicz. 1977. Topics in phonological theory. New York, NY: Academic Press.

Korsah, Sampson. 2015. Ga imperatives and richness of the base. Universität Leipzig, manuscript.

Kruspe, Nicole. 2004. A grammar of Semelai. Cambridge: Cambridge University Press.
Kügler, Frank. 2015. Phonological phrasing and ATR vowel harmony in Akan. Phonology 32(1). 177-204.

Kutsch Lojenga, Constance. 1994. Ngiti: a central sudanic language of Zaire. Köln: Rüdiger Köppe.

Kutsch-Lojenga, Constance. 2008. Nine vowels and ATR vowel harmony in Lika, a Bantu language in DR Congo. Africana Linguistica 14(1). 63-84.
de Lacy, Paul. 2007. The interaction of tone, sonority, and prosodic structure. In Paul de Lacy (ed.), The Cambridge handbook of phonology, 281-308. Cambridge: Cambridge University Press.

Lapowila, Hans. 1981. A generative approach to the phonology of Bahasa Indonesia. Canberra: The Australian National University.

Leben, William. 1973. Suprasegmental phonology. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Leben, William. 1978. The representation of tone. In Victoria A Fromkin (ed.), Tone: A linguistic survey, 177-219. New York: Academic Press.

Lees, Robert B. 1966. On the interpretation of a Turkish vowel alternation. Anthropological Linguistics 8(9). 32-39.

Li, Paul Jen-kuei \& Shigeru Tsuchida. 2006. Kavalan dictionary. Taipei: Academia Sinica.
Lieber, Rochelle. 1987. An integrated theory of autosegmental processes. Albany, NY: SUNY Press.

Lohmann, Arne. 2017. Phonological properties of word classes and directionality in conversion. Word Structure 10(2). 204-234.

Lohmann, Arne. 2018. Cut (n) and cut (v) are not homophones: Lemma frequency affects the duration of noun-verb conversion pairs. Journal of Linguistics 54(4). 753-777.

Lohmann, Arne. 2020a. No acoustic correlates of grammatical class: a critical reexamination of Sereno and Jongman (1995). Phonetica 77(6). 429-440.

Lohmann, Arne. 2020b. Nouns and verbs in the speech signal: Are there phonetic correlates of grammatical category? Linguistics 58(6). 1877-1911.

Lohmann, Arne \& Erin Conwell. 2020. Phonetic effects of grammatical category: How category-specific prosodic phrasing and lexical frequency impact the duration of nouns and verbs. Journal of Phonetics 78.

Loughnane, Robyn. 2009. A grammar of Oksapmin. Melbourne: University of Melbourne dissertation.

Lovick, Olga. 2020. A grammar of Upper Tanana, volume 1: Phonology, lexical classes, morphology. Lincoln, NE: University of Nebraska Press.

Lubowicz, Anna. 2002. Derived Environment Effects in Optimality Theory. Lingua 112(4). 243-280.

Lubowicz, Anna. 2003. Local conjunction and comparative markedness. Theoretical Linguistics 29(1-2). 101-112.

Lubowicz, Anna. 2010. Infixation as morpheme-absorption. In Steve Parker (ed.), Phonological argumentation: Essays on evidence and motivation, 261-284. London: Equinox.

Maganga, Clement \& Thilo Schadeberg. 1992. Kinyamwezi: grammar, texts, vocabulary. Köln: Köppe.

Magri, Giorgio. 2018. Idempotency in Optimality Theory. Journal of Linguistics 54(1). 139-187.

Mahanta, Shakuntala. 2008. Directionality and locality in vowel harmony: With special reference to vowel harmony in Assamese. Utrecht: Utrecht Institute of Linguistics dissertation.

Mahanta, Shakuntala. 2012. Locality in exceptions and derived environments in vowel harmony. Natural Language \& Linguistic Theory 30(4). 1109-1146.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. University of Pennsylvania working papers in linguistics 4(2).

Marlo, Michael. 2013. Verb tone in Bantu languages: micro-typological patterns and research methods. Africana Linguistica 19(1). 137-234.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Cambridge, MA: MIT dissertation.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. Linguistic Inquiry 12(3). 373-418.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. Linguistic Inquiry 17(2). 207-263.

McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. In M. Catherine Gruber, Derrick Higgins, Kenneth Olson \& Tamra Wysocki (eds.), CLS 32, part 2: The panels, 123-150. Chicago, IL: Chicago Linguistic Society.

McCarthy, John J. 2000. Harmonic serialism and parallelism. In Proceedings of the north east linguistics society 30, vol. 2, .

McCarthy, John J. 2004. Headed spans and autosegmental spreading. Linguistics Department Faculty Publication Series 42.

McCarthy, John J. 2012. Pausal phonology and morpheme realization. In Toni Borowsky, Shigeto Kawahara, Takahito Shinya \& Mariko Sugahara (eds.), Prosody matters: Essays in honor of Lisa Selkirk, 341-373. London: Equinox.

McCarthy, John J. \& Alan Prince. 1993. Generalized alignment. In Yearbook of morphology 1993, 79-153. Springer.

McCarthy, John J. \& Alan Prince. 1995. Faithfulness and reduplicative identity. University of Massachusetts Occasional Papers in Linguistics 18. 249-384.

McGregor, William. 1990. A functional grammar of Gooniyandi. Amsterdam: Benjamins.
McPherson, Laura. 2013. A grammar of Tommo So. Berlin: De Gruyter Mouton.
McPherson, Laura \& Matthew S. Dryer. 2021. The tone system of Poko-Rawo (Skou). Phonological Data and Analysis 3(1). 1-32.

Meira, Sérgio. 1999. A grammar of Tiriyó. Houston, TX: Rice University dissertation.
Mellese, Gelaneh Alemu. 2017. Documentation and grammatical description of Tapo. Addis Ababa: Addis Ababa University dissertation.

Mester, Armin. 1987. Studies in tier structure. Amherst, MA: University of Massachusetts dissertation.

Mielke, Jeff. 2008. The emergence of distinctive features. Oxford: Oxford University Press.
Miller, Taylor L \& Hannah Sande. 2021. Is word-level recursion actually recursion? Languages 6(2).

Miller, Wick R. 1965. Acoma grammar and texts. Berkeley, CA: University of California Press.

Minter, Paul. 2009. Iyo grammar sketch. Ukarumpa: SIL-PNG.

Modi, Yankee. 2017. The Milang language. Bern: University of Bern dissertation.
Mohanan, Tara. 1989. Syllable structure in Malayalam. Linguistic Inquiry 589-625.
Myers, Scott. 1997. OCP effects in Optimality Theory. Natural Language \& Linguistic Theory 15(4). 847-892.

Nespor, Marina \& Irene Vogel. 2012. Prosodic phonology. Berlin: De Gruyter Mouton.
Newman, Paul. 1986. Tone and affixation in Hausa. Studies in African linguistics 17(3). 249-268.

Ngunga, Armindo Saúl Atelela. 1997. Lexical phonology and morphology of the Ciyao verb stem. Berkeley, CA: University of California dissertation.

Noonan, Michael. 1992. A grammar of Lango. Berlin: Walter de Gruyter.
Nordlinger, Rachel. 1998. A grammar of Wambaya, Northern Territory (Australia). Canberra: Pacific Linguistics.

Odden, David. 1981. Problems in tone assignment in Shona. Urbana-Champaign, IL: University of Illinois dissertation.

Odden, David. 1984. Stem tone assignment in Shona. In George N. Clements \& J. Goldsmith (eds.), Autosegmental studies in Bantu tone, 255-280. Dodrecht: Foris.

O'Grady, Geoffrey Noel. 1963. Nyangumata grammar. Bloomington, IN: Indiana University dissertation.

O’Keefe, Michael. 2007. Transparency in span theory. In Leah Bateman, Adam Werle, Michael O'Keefe \& Ehren Reilly (eds.), University of massachusetts occasional papers in linguistics, vol. 33, 239-258. Amherst: GLSA.

Onziga, Yuga Juma \& Leoma Gilley. 2012. Phonology of Kakuwâ (Kakwa). Occasional Papers in the Study of Sudanese Languages 10. 1-15.
van Oostendorp, Marc. 1999. Italian s-voicing and the structure of the phonological word. In S.J. Hannahs \& Mike Davenport (eds.), Issues in phonological structure, 195-212. Amsterdam: John Benjamins.
van Oostendorp, Marc. 2001. Back to the base: On the richness of the base hypothesis. In Marc van Oostendorp \& Elena Anagnostopoulou (eds.), Progress in grammar: articles at the 20th anniversary of the comparison of grammatical models group in Tilburg, Utrecht: Roccade.
van Oostendorp, Marc. 2007. Derived environment effects and consistency of exponence. In Silvya Blaho, Patrik Bye \& Martin Krämer (eds.), Freedom of analysis?, 123-148. Berlin: Mouton de Gruyter.
van Oostendorp, Marc. 2014. Selective lexicon optimization. Lingua 142. 76-84.
Orgun, Cemil Orhan. 1996. Sign-based morphology and phonology with special attention to Optimality Theory. Berkeley, CA: University of California dissertation.

Orgun, Cemil Orhan \& Ronald L Sprouse. 1999. From "MParse" to "Control": Deriving ungrammaticality. Phonology 16(2). 191-224.

Orgun, Orhan \& Ronald Sprouse. 2008. A freer input: Yowlumne opacity and the enriched input model. In Sylvia Blaho, Patrik Bye \& Martin Krämer (eds.), Freedom of analysis?, 93-122. Berlin: De Gruyter Mouton.

Orie, Olanike Ola. 2003. Two harmony theories and high vowel patterns in Ebira and Yoruba. The Linguistic Review 20(1). 1-35.

Otero, Manuel. 2015. [+ATR] dominant vowel harmony except when it's not? Evidence from Ethiopian Komo. In Ruth Kramer, Elizabeth Zsiga \& One Tlale Boyer (eds.), 44 th Annual Conference on African Linguistics, 212-220. Somerville: Cascadilla Proceedings Project.

Ourso, Meterwa A. 1989. Phonological processes in the noun class system of Lama. Studies in African linguistics 20(2). 151-178.

Paster, Mary. 2000. Issues in the tonology of Gã. Columbus, OH: Ohio State University dissertation. Honor's Thesis.

Pater, Joe. 2001. Austronesian nasal substitution revisited. In Linda Lombardi (ed.), Segmental phonology in optimality theory: Constraints and representations, 159-182. Cambridge: Cambridge University Press.

Pater, Joe. 2007. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. University of Massachusetts Occasional Papers in Linguistics 32.

Peacock, Wesley. 2007. The phonology of Nkonya. Legon: University of Ghana.

Peperkamp, Sharon. 1996. On the prosodic representation of clitics. In Ursula Kleinherz (ed.), Interfaces in phonology, 104-128. Berlin: Akademie Verlag.

Plank, Frans \& Elena Filimonova. 2000. The universals archive: A brief introduction for prospective users. Language Typology and Universals 53(1). 109-123.

Prince, Alan. 1998. Two lectures on Optimality Theory. Handout from Phonology Forum 1998, Kobe University.

Prince, Alan \& Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Available online at http://roa.rutgers.edu/files/537-0802/ 537-0802-PRINCE-0-0.PDF.

Pulleyblank, Douglas. 1986. Tone in lexical phonology. Cambridge, MA: MIT dissertation.
Raffelsiefen, Renate. 2000. Evidence for word-internal phonological words in German. In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop \& Oliver Teuber (eds.), Deutsche grammatik in theorie und praxis, 43-56. Tübingen: Niemeyer.

Rai, Netra Mani. 2016. A grammar of Dumi. Kirtipur: Tribhuvan University dissertation.
Rarrick, Samantha Carol. 2017. A tonal grammar of Kere (Papuan) in typological perspective. Manoa: University of Hawai'i dissertation.

Rasin, Ezer. 2016. Morpheme structure constraints and blocking in nonderived environments. Manuscript.

Rasin, Ezer \& Roni Katzir. 2020. A conditional learnability argument for constraints on underlying representations. Journal of Linguistics 56(4). 745-773.

Reiner, Erica. 1965. A linguistic analysis of Akkadian. The Hague: Mouton \& Co.
Revithiadou, Anthi. 2007. Colored turbid accents and containment: A case study from lexical stress. In Silvya Blaho, Patrik Bye \& Martin Krämer (eds.), Freedom of analysis?, 149-174. Berlin: Mouton de Gruyter.

Rolle, Nicholas, Florian Lionnet \& Matthew Faytak. 2020. Areal patterns in the vowel systems of the Macro-Sudan Belt. Linguistic Typology 24(1). 113-179.

Rolle, Nicholas Revett. 2018. Grammatical tone: Typology and theory. Berkeley: University of California.

Rose, Sharon \& Rachel Walker. 2004. A typology of consonant agreement as correspondence. Language 80(3). 475-531.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. Phonology Yearbook 3. 371-405.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle \& Alan Yu (eds.), The handbook of phonological theory, 435-483. Oxford: Blackwell.

Shay, Erin. 2021. A grammar of Giziga: A Chadic language of Far North Cameroon. Leiden: Brill.

Shih, Stephanie \& Sharon Inkelas. 2018. Autosegmental aims in surface-optimizing phonology. Linguistic Inquiry 50(1). 137-196.

Silva, Wilson. 2012. A descriptive grammar of Desano. Salt Lake City, UT: The University of Utah dissertation.

Siptár, Péter \& Miklós Törkenczy. 2000. The phonology of Hungarian. Oxford: Oxford University.

Smeets, Ineke. 2008. A grammar of Mapuche. Berlin: De Gruyter Mouton.
Smith, Jennifer L. 2002. Phonological augmentation in prominent positions. Amherst: University of Massachusetts dissertation.

Smith, Jennifer L. 2011. Category-specific effects. In The blackwell companion to phonology, 2439-2463. Oxford: Wiley Blackwell.

Smith, Jennifer L. 2018. Feature change is not like deletion: Saltation in harmonic grammar. talk, given at LSA 2018.

Smolensky, Paul et al. 2006. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In Geraldine Legendre Paul Smolensky (ed.), The harmonic mind: From neural computation to optimalitytheoretic grammar, vol. 2, 27-160. Cambridge: MIT Press.

Stanley, Richard. 1967. Redundancy rules in phonology. Language 43(2). 393-436.
Stanton, Juliet. 2020. Aggressive reduplication and dissimilation in Sundanese. Phonological Data and Analysis 2(5). 1-35.

Staroverov, Peter. 2020. Opaque distributional generalisations in Tundra Nenets. Phonology 37(2). 297-328.

Stewart, John M. 1983. Akan vowel harmony: the word structure conditions and the floating vowels. Studies in African linguistics 14(2). 111-140.

Stirtz, Timothy. 2014. Mundari phonology. Dallas: SIL.
Suzuki, Keiichiro. 1998. A typological investigation of dissimilation. Tucson, AZ: University of Arizona dissertation.

Tallman, Adam J. 2018. A grammar of Chácobo, a southern pano language of the northern Bolivian Amazon. Austin: University of Texas dissertation.

Tebay, Sören E. \& Eva Zimmermann. 2020. Exceptionality in Assamese vowel harmony: A phonological account. Glossa: a journal of general linguistics 5(1). 1-49.

Temsunungsang, T. 2009. Aspects of the prosodic phonology of Ao: An inter-dialectal study. Hyderabad: The English and Foreign Languages University dissertation.

Temsunungsang, T. 2017. Tonal alternations in Chungli (Ao): A case of multiple prosodic words. In Fabric of Indian linguistics: A festschrift in honour of Prof. Udaya Narayana Singh, 132-150. New Delhi: Lakshi.

Teng, Stacy. 2008. A reference grammar of Puyuma, an Austronesian language of Taiwan. Canberra: Pacific Linguistics.

Tesar, Bruce. 2013. Output-driven phonology: Theory and learning. Cambridge: Cambridge University Press.

Topintzi, Nina. 2003. Issues of locality and morphologically induced non-identity in the N. Karanga assertive and non-assertive patterns. UCL Working Papers in Linguistics 15. 327-354.

Torres-Tamarit, Francesc. 2016. Compensatory and opaque vowel lengthening in harmonic serialism. In Harmonic grammar and harmonic serialism, Sheffield: Equinox.

Torres-Tamarit, Francesc, Kathrin Linke \& Maria del Mar Vanrell. 2017. Opacity in Campidanian Sardinian metaphony. Natural Language \& Linguistic Theory 35(2). 549-576.

Trommer, Jochen. 2011. Phonological aspects of Western Nilotic mutation morphology. Habilitation.

Ussishkin, Adam. 2005. A fixed prosodic theory of nonconcatenative templatic morphology. Natural Language छ Linguistic Theory 23(1). 169-218.

Vaysman, Olga. 2002. Against richness of the base: Evidence from Nganasan. In Julie Larson \& Mary Paster (eds.), Annual Meeting of the Berkeley Linguistics Society, vol. 28 1, 327-338.

Vaysman, Olga. 2009. Segmental alternations and metrical theory. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Villard, Stéphanie. 2015. The phonology and morphology of Zacatepec Eastern Chatino. Austin, TX: University of Texas dissertation.

Voinov, Vitaly. 2016. Sibilant harmony in Tuvan roots. Rodnoj jazyk 1(4). 8-20.

Walker, Rachel. 2005. Weak triggers in vowel harmony. Natural Language \& Linguistic Theory 23(4). 917-989.

Walker, Rachel. 2010. Nonmyopic harmony and the nature of derivations. Linguistic Inquiry 41(1). 169-179.

Walker, Rachel. 2011. Vowel patterns in language. Cambridge: Cambridge University Press.

Waters, Bruce. 1989. Djinang and Djinba: A grammatical and historical perspective. Canberra: Pacific Linguistics.

Watters, John Robert. 1981. A phonology and morphology of Ejagham-with notes on dialect variation. Los Angeles, CA: University of California dissertation.

Weber, David. 1989. A grammar of Huallaga (Huanuco) Quechua. Berkeley: University of California Press.

Wilson, Jennifer. 2017. A grammar of Yeri: a Torricelli language of Papua New Guinea. Buffalo, NY: State University of New York dissertation.

Yip, Moira. 1988. Template morphology and the direction of association. Natural Language \& Linguistic Theory 6(4). 551-577.

Yu, Alan. 2004. Infixing with a vengeance: Pingding Mandarin infixation. Journal of East Asian Linguistics 13(1). 39-58.

Yu, Alan. 2007. A natural history of infixation. Oxford: Oxford University Press.
Zec, Draga. 1994. Sonority constraints on prosodic structure. New York: Garland, NY.
Zimmer, Karl E. 1969. Psychological correlates of some .Turkish morpheme structure conditions. Language 309-321.

Zimmerman, Johannes. 1858. A grammatical sketch and vocabulary of the Akra - or Ga-language with an appendix on the Adanme-dialect. Stuttgart: J.F. Steinkopf.

Zoll, Cheryl. 2003. Optimal tone mapping. Linguistic Inquiry 34(2). 225-268.
Zuraw, Kie \& Yu-An Lu. 2009. Diverse repairs for multiple labial consonants. Natural Language \& Linguistic Theory 27(1). 197-224.

## Appendix: DoCoMD

In this appendix, all information from the database on constraints on monomorphemic domains is included as well as the samples used in chapters 2, 3, and 4. All tables are sorted by language name. Table 1 lists all languages used in DoCoMD and information associated with them. Table 2 gives all information on languages used in the g-sample in chapter 2. Table 3 includes a standardized description of each COMD as well as a quotation from the original source. References found in the quotation have been excluded to ease reading. The constraint properties of all entries in DoCoMD are given in table 4. Here and in the following, entries in different tables can be linked by their index. The information on interaction of CoMDs with infixation and vowel harmony is given in table 5 and 6, respectively. Finally, sources used in DoCoMD are listed in table 7, including an indication of the entries that are also present in P-Base (Mielke 2008). For a more detailed description of the format of DoCoMD see chapter 2.
Table 1: List of languages in DoCoMD

| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| tbw | Aborlan Tagbanwa | tagb1258 | Austronesian | Greater Central Philippine | Papuanesia |
| kjq | Acoma | acom1246 | Keresan | Keresan | North America |
| aha | Ahanta | ahan1243 | Atlantic-Congo | Kwa | Africa |
| aka | Akan | akan1250 | Atlantic-Congo | Kwa | Africa |
| akk | Akkadian | akka1240 | Afro-Asiatic | Semitic | Eurasia |
| yor | Akure Yoruba | ondo1239 | Atlantic-Congo | Defoid | Africa |
| rwm | Amba | amba1263 | Atlantic-Congo | Bantoid | Africa |
| gwx | Anum | guaa1238 | Atlantic-Congo | Kwa | Africa |
| arb | Arabic | stan1318 | Afro-Asiatic | Semitic | Eurasia |
| asm | Assamese | assa1263 | Indo-European | Indic | Eurasia |
| avi | Avikam | avik1243 | Atlantic-Congo | Kwa | Africa |
| avu | Avokaya | avok1242 | Central Sudanic | Moru-Ma'di | Africa |
| bdh | Baka | baka1274 | Central Sudanic | Bongo-Bagirmi | Africa |
| bct | Bendi | bend1260 | Central Sudanic | Mangbutu-Efe | Africa |
| bdi | Burun | buru1301 | Nilotic | Nilotic | Africa |
| sro | Campidanese Sardinian | camp1261 | Indo-European | Romance | Eurasia |
| yue | Cantonese | cant1236 | Sino-Tibetan | Chinese | Eurasia |
| cao | Chácobo | chac1251 | Pano-Tacanan | Panoan | South America |
| ncu | Chumburung | chum1261 | Atlantic-Congo | Kwa | Africa |
| njo | Chungli Ao | chon1286 | Sino-Tibetan | Kuki-Chin | Eurasia |
| yao | Ciyao | yaoo1241 | Atlantic-Congo | Bantoid | Africa |
| com | Comanche | coma1245 | Uto-Aztecan | Numic | North America |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| tfn | Dena'ina | tana1289 | Athabaskan-Eyak-Tlingit | Athapaskan | Nucanoan |
| des | Desano | desa1247 | Tucanoan | Surmic | America |
| did | Didinga | didi1258 | Surmic | Western Pama-Nyungan | Africa |
| dji | Djinang | djin1253 | Pama-Nyungan | Australia |  |
| dus | Dumi | dumi1241 | Sino-Tibetan | Mahakiranti | Eurasia |
| nld | Dutch | dutc1256 | Indo-European | Germanic | Eurasia |
| mhr | Eastern Mari | east2328 | Uralic | Mari | Eurasia |
| etu | Ejagham | ejag1239 | Atlantic-Congo | Bantoid | Africa |
| yor | Ekiti Yoruba | ekit1244 | Atlantic-Congo | Defoid | Africa |
| eng | English | stan1293 | Indo-European | Germanic | Eurasia |
| fmp | Fe'Fe'-Bamileke | fefe1239 | Atlantic-Congo | Bantoid | Africa |
| fvr | Fur | furr1244 | Furan | Africa |  |
| gaa | Gã | gaaa1244 | Atlantic-Congo | Kwa | Africa |
| cab | Garifuna | gari1256 | Arawakan | Caribbean Arawakan | North America |
| deu | German | stan1295 | Indo-European | Germanic | Eurasia |
| acd | Gichode | giky1238 | Atlantic-Congo | Kwa | Africa |
| toh | Gitonga | gito1238 | Atlantic-Congo | Bantoid | Africa |
| gis | Giziga | north3047 | Afro-Asiatic | Biu-Mandara | Africa |
| gjn | Gonja | gonj1241 | Atlantic-Congo | Kwa | Africa |
| gni | Gooniyandi | goon1238 | Bunaban | Bunaban | Australia |
| rub | Gungu | gung1250 | Atlantic-Congo | Bantoid | Africa |
| heb | Hebrew | hebr1245 | Afro-Asiatic | Semitic | Eurasia |
| qub | Huallaga Huánuco Quechua | hual1241 | Quechuan | Quechuan | South America |
| hun | Hungarian | hung1274 | Uralic | Ugric | Eurasia |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| huz | Hunzib | hunz1247 | Nakh-Dagestanian | Avar-Andic-Tsezic | Eurasia |
| yor | Ifaki Yoruba | ekit1244 | Atlantic-Congo | Defoid | Africa |
| yor | Ijesa Yoruba | ijes1238 | Atlantic-Congo | Defoid | Africa |
| ind | Indonesian | indo1316 | Austronesian | Malayo-Sumbawan | Papuanesia |
| yor | Irun Yoruba | ondo1239 | Atlantic-Congo | Defoid | Africa |
| ita | Italian | ital1282 | Indo-European | Romance | Eurasia |
| nca | Iyo | iyoo1238 | Nuclear Trans New Guinea | Finisterre-Huon | Papuanesia |
| jav | Javanese | java1254 | Austronesian | Javanese | Papuanesia |
| jum | Jumjum | jumj1238 | Nilotic | Nilotic | Africa |
| ksp | Kaba | kaba1281 | Central Sudanic | Bongo-Bagirmi | Africa |
| keo | Kakwa | kakw1240 | Nilotic | Nilotic | Africa |
| kis | Kalasha | kala1372 | Indo-European | Indic | Eurasia |
| xnb | Kanakanavu | kana1286 | Austronesian | Psouic | Papuanesia |
| pam | Kapampangan | pamp1243 | Austronesian | Central Luzon | Papuanesia |
| ckv | Kavalan | kava1241 | Austronesian | Papuanesia |  |
| sst | Kere | kere1300 | Nuclear Trans New Guinea | Chimbu | Papuanesia |
| kzj | Kimaragang | coas1294 | Austronesian | North Borneo | Africa |
| nnb | Kinande | nand1264 | Atlantic-Congo | Bantoid | Africa |
| nym | Kinyamwezi | nyam1276 | Atlantic-Congo | Bantoid | Africa |
| xom | Komo | komo1258 | Koman | Koman | Eurasia |
| kpy | Koryak | kory1246 | Chukotko-Kamchatkan | Northern Chukotko-Kamchatkan | Africa |
| kye | Krachi | krac1238 | Atlantic-Congo | Kwa | Africa |
| bfa | Kuku | kuku1285 | Nilotic | Nilotic | North America |
| lkt | Lakhota | lako1247 | Siouan | Core Siouan |  |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| las | Lama | lama1275 | Atlantic-Congo | Gur | Africa |
| lag | Langi | lang1320 | Atlantic-Congo | Bantoid | Nilotic |
| laj | Lango | lang1324 | Nilotic | Mangbutu-Efe | Africa |
| les | Lese | lese1243 | Central Sudanic | Bodic | Africa |
| bod | Lhasa Tibetan | utsa1239 | Sino-Tibetan | Northern Luzon | Eurasia |
| kmk | Limos Kalinga | limo1248 | Austronesian | Greater Central Philippine | Papuanesia |
| llq | Lolak | lola1250 | Austronesian | Papuanesia |  |
| dni | Lower Grand Valley Dani | lowe1415 | Nuclear Trans New Guinea | Dani | Papuanesia |
| lgg | Lugbara | lugb1240 | Central Sudanic | Moru-Ma'di | Africa |
| mal | Malayalam | mala1464 | Dravidian | South Dravidian | Eurasia |
| mgq | Malila | mali1279 | Atlantic-Congo | Bantoid | Africa |
| mdk | Mangbutu | mang1396 | Central Sudanic | Mangbutu-Efe | Africa |
| arn | Mapuche | mapu1245 | Araucanian | Araucanian | South America |
| vma | Martuthunira | mart1255 | Pama-Nyungan | Western Pama-Nyungan | Australia |
| mdm | Mayogo | mayo1261 | Atlantic-Congo | Ubangi | Africa |
| tay | Mayrinax Atayal | mayr1237 | Austronesian | Atayalic | Papuanesia |
| jkr | Milang | mila1245 | Sino-Tibetan | Tani | Eurasia |
| mnb | Muna | muna1247 | Austronesian | Celebic | Papuanesia |
| mqu | Mundari | mand1425 | Nilotic | Nilotic | Africa |
| sur | Mupun | mupu1234 | Afro-Asiatic | West Chadic | Africa |
| acx | Muscat Arabic | oman1239 | Afro-Asiatic | Semitic | Eurasia |
| mxh | Mvuba | mvub1239 | Central Sudanic | Mangbutu-Efe | Africa |
| jup | Naduhup | hupd1244 | Naduhup | South America |  |
| naw | Nawuri | nawup | Africa |  |  |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| niy | Ngiti | ngit1239 | Central Sudanic | Lendu | Africa |
| mks | Nieves Mixtec | sila1250 | Otomanguean | Mixtecan | Kwa |
| nko | Nkonya | nkon1248 | Atlantic-Congo | Malayo-Sumbawan | Africa |
| ace | Northern Acehnese | achi1257 | Austronesian | Papuanesia |  |
| nso | Northern Sotho | nort3233 | Atlantic-Congo | Bantoid | Africa |
| nna | Nyangumarta | nyan1301 | Pama-Nyungan | Western Pama-Nyungan | Australia |
| nzb | Nzebi | njeb1242 | Atlantic-Congo | Bantoid | Africa |
| opm | Oksapmin | oksa1245 | Nuclear Trans New Guinea | Oksapmin | Papuanesia |
| lgn | Opouuo | opuu1239 | Koman | Koman | Africa |
| pwn | Paiwan | paiw1248 | Austronesian | Paiwan | Papuanesia |
| pau | Palauan | pala1344 | Austronesian | Palauan | Papuanesia |
| lkr | Päri | ari1256 | Nilotic | Africa |  |
| spa | Pasiego Spanish | stan1288 | Indo-European | Romance | Eurasia |
| cmn | Pingding Mandarin | mand1415 | Sino-Tibetan | Chinese | Eurasia |
| fuc | Pulaar | pula1263 | Atlantic-Congo | Peul-Serer | Africa |
| puu | Punu | punu1239 | Atlantic-Congo | Bantoid | Africa |
| pyu | Puyuma | puyu1239 | Austronesian | Puyuma | Papuanesia |
| rus | Russian | russ1263 | Indo-European | Slavic | Eurasia |
| bps | Sarangani Blaan | sara1326 | Austronesian | Billic | Papuanesia |
| mbs | Sarangani Manobo | sara1327 | Austronesian | Greater Central Philippine | Papuanesia |
| trv | Seediq | taro1264 | Austronesian | Atayalic | Papuanesia |
| sza | Semelai | Aslian | Eurasia |  |  |
| sot | Sesotho | Austroasiatic | Bantoid | Africa |  |
| tsn | Setswana | Bantoid | Africa |  |  |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| kca | Southwestern Khanty | khan1279 | Uralic | Ugric | Eurasia |
| erg | Sye | sie1239 | Austronesian | Oceanic | Papuanesia |
| tgl | Tagalog | taga1270 | Austronesian | Greater Central Philippine | Papuanesia |
| tlj | Talinga Bwisi | tali1258 | Atlantic-Congo | Bantoid | Africa |
| tpj | Tapieté | tapi1253 | Tupian | Tupi-Guaraní | South America |
| ssf | Thao | thao1240 | Austronesian | Western Plains Austronesian | Papuanesia |
| tih | Timugon Murut | timu1262 | Austronesian | North Borneo | Papuanesia |
| suq | Tirmaga | suri1267 | Surmic | Surmic | Africa |
| bbc | Toba Batak | bata1289 | Austronesian | Northwest Sumatra-Barrier Islands | Papuanesia |
| dto | Tommo So Dogon | tomm1242 | Dogon | Aogon | Africa |
| tri | Trió | trio1238 | Cariban | Cariban | South America |
| tso | Tsonga | tson1249 | Atlantic-Congo | Bantoid | Africa |
| tur | Turkish | nucl1301 | Turkic | Turkic | Eurasia |
| ulw | Ulwa | ulwa1239 | Misumalpan | Misumalpan | North America |
| uur | Ura | urav1235 | Austronesian | Oceanic | Papuanesia |
| ghl | Uyncu | ghul1238 | Nubian | Nubian | Africa |
| wmb | Wambaya | nucl1328 | Mirndi | Wambayan | Australia |
| yak | Yakima Sahaptin | yaki1237 | Sahaptian | Sahaptin | North America |
| yev | Yeri | yapu1240 | Nuclear Torricelli | Wapei-Palei | Papuanesia |
| ctz | Zacatepec Chatino | zaca1242 | Otomanguean | Ubangi | North America |
| zne | Zande | zand1248 | Atlantic-Congo | Bantoid | Africa |
| zul | Zulu | zulu1248 | Atlantic-Congo |  | Africa |

Table 2: List of languages in g-sample

| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| kjq | Acoma | acom1246 | Keresan | Keresan | North America |
| aha | Ahanta | ahan1243 | Atlantic-Congo | Kwa | Africa |
| aka | Akan | akan1250 | Atlantic-Congo | Kwa | Africa |
| akk | Akkadian | akka1240 | Afro-Asiatic | Semitic | Eurasia |
| rwm | Amba | amba1263 | Atlantic-Congo | Bantoid | Africa |
| gwx | Anum | guaa1238 | Atlantic-Congo | Kwa | Africa |
| arb | Arabic | stan1318 | Afro-Asiatic | Semitic | Eurasia |
| asm | Assamese | assa1263 | Indo-European | Indic | Eurasia |
| avi | Avikam | avik1243 | Atlantic-Congo | Kwa | Africa |
| avu | Avokaya | avok1242 | Central Sudanic | Moru-Ma'di | Africa |
| bdh | Baka | baka1274 | Central Sudanic | Bongo-Bagirmi | Africa |
| bdi | Burun | buru1301 | Nilotic | Ailotic | Africa |
| sro | Campidanese Sardinian | camp1261 | Indo-European | Romance | Eurasia |
| yue | Cantonese | cant1236 | Sino-Tibetan | Chinese | Eurasia |
| cao | Chácobo | chac1251 | Pano-Tacanan | Panoan | South America |
| ncu | Chumburung | chum1261 | Atlantic-Congo | Kwa | Africa |
| njo | Chungli Ao | chon1286 | Sino-Tibetan | Kuki-Chin | Eurasia |
| yao | Ciyao | yaoo1241 | Atlantic-Congo | Bantoid | Africa |
| com | Comanche | coma1245 | Uto-Aztecan | Numic | North America |
| tfn | Dena'ina | tana1289 | Athabaskan-Eyak-Tlingit | Athapaskan | Nucanoan |
| des | Desano | desa1247 | Tucanoan | Surmic | South America |
| did | Didinga | didi1258 | Surmic | Africa |  |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| dji | Djinang | djin1253 | Pama-Nyungan | Western Pama-Nyungan | Australia |
| dus | Dumi | dumi1241 | Sino-Tibetan | Mahakiranti | Eurasia |
| nld | Dutch | dutc1256 | Indo-European | Germanic | Marasia |
| mhr | Eastern Mari | east2328 | Uralic | Maric | Eurasia |
| etu | Ejagham | ejag1239 | Atlantic-Congo | Bantoid | Africa |
| eng | English | stan1293 | Indo-European | Germanic | Eurasia |
| fmp | Fe'Fe'-Bamileke | fefe1239 | Atlantic-Congo | Bantoid | Africa |
| fvr | Fur | furr1244 | Furan | Fur | Africa |
| gaa | Gã | gaaa1244 | Atlantic-Congo | Kwa | Africa |
| cab | Garifuna | gari1256 | Arawakan | Caribbean Arawakan | North America |
| deu | German | stan1295 | Indo-European | Germanic | Eurasia |
| toh | Gitonga | gito1238 | Atlantic-Congo | Bantoid | Africa |
| gis | Giziga | north3047 | Afro-Asiatic | Biu-Mandara | Africa |
| gjn | Gonja | gonj1241 | Atlantic-Congo | Kwa | Africa |
| gni | Gooniyandi | goon1238 | Bunaban | Australia |  |
| rub | Gungu | gung1250 | Atlantic-Congo | Bantoid | Africa |
| heb | Hebrew | hebr1245 | Afro-Asiatic | Semitic | Eurasia |
| qub | Huallaga Huánuco Quechua | hual1241 | Quechuan | Quechuan | South America |
| hun | Hungarian | hung1274 | Uralic | Ugric | Eurasia |
| huz | Hunzib | hunz1247 | Nakh-Dagestanian | Avar-Andic-Tsezic | Eurasia |
| yor | Ijesa Yoruba | ijes1238 | Atlantic-Congo | Defoid | Africa |
| ind | Indonesian | indo1316 | Austronesian | Malayo-Sumbawan | Papuanesia |
| ita | Italian | ital1282 | Indo-European | Romance | Eurasia |
| nca | Iyo | iyoo1238 | Nuclear Trans New Guinea | Finisterre-Huon | Papuanesia |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| jav | Javanese | java1254 | Austronesian | Javanese | Nilotic |
| jum | Jumjum | jumj1238 | Nilotic | Nilotic | Papuanesia |
| keo | Kakwa | kakw1240 | Nilotic | Andic | Africa |
| kis | Kalasha | kala1372 | Indo-European | Tsouic | Eurasia |
| xnb | Kanakanavu | kana1286 | Austronesian | Central Luzon | Papuanesia |
| pam | Kapampangan | pamp1243 | Austronesian | East Formosan | Papuanesia |
| ckv | Kavalan | kava1241 | Austronesian | Papuanesia |  |
| sst | Kere | kere1300 | Nuclear Trans New Guinea | Chimbu | Papuanesia |
| kzj | Kimaragang | coas1294 | Austronesian | North Borneo | Papuanesia |
| nym | Kinyamwezi | nyam1276 | Atlantic-Congo | Bantoid | Africa |
| xom | Komo | komo1258 | Koman | Koman | Africa |
| kpy | Koryak | kory1246 | Chukotko-Kamchatkan | Northern Chukotko-Kamchatkan | Eurasia |
| bfa | Kuku | kuku1285 | Nilotic | Nilotic | Africa |
| lkt | Lakhota | lako1247 | Siouan | Core Siouan | Aurth America |
| las | Lama | lama1275 | Atlantic-Congo | Bantoid | Africa |
| lag | Langi | $l a n g 1320$ | Atlantic-Congo | Nilotic | Africa |
| laj | Lango | lang1324 | Nilotic | Bodic | Africa |
| bod | Lhasa Tibetan | utsa1239 | Sino-Tibetan | Northern Luzon | Eurasia |
| kmk | Limos Kalinga | limo1248 | Austronesian | Papuanesia |  |
| dni | Lower Grand Valley Dani | lowe1415 | Nuclear Trans New Guinea | Dani | Papuanesia |
| mal | Malayalam | mala1464 | Dravidian | South Dravidian | Bantoid |
| mgq | Malila | mali1279 | Atlantic-Congo | Araucanian | Western Pama-Nyungan |
| arn | Mapuche | mapu1245 | Araucanian | Africa |  |
| vma | Martuthunira | mart1255 | Pama-Nyungan | Australia |  |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| tay | Mayrinax Atayal | mayr1237 | Austronesian | Atayalic | Papuanesia |
| jkr | Milang | mila1245 | Sino-Tibetan | Tani | Eurasia |
| mnb | Muna | muna1247 | Austronesian | Celebic | Papuanesia |
| mqu | Mundari | mand1425 | Nilotic | Africa |  |
| sur | Mupun | mupu1234 | Afro-Asiatic | West Chadic | Africa |
| acx | Muscat Arabic | oman1239 | Afro-Asiatic | Semitic | Eurasia |
| jup | Naduhup | hupd1244 | Naduhup | Nadahup | South America |
| niy | Ngiti | ngit1239 | Central Sudanic | Lendu | Africa |
| mks | Nieves Mixtec | sila1250 | Otomanguean | Mixtecan | North America |
| nko | Nkonya | nkon1248 | Atlantic-Congo | Kwa | Africa |
| ace | Northern Acehnese | achi1257 | Austronesian | Malayo-Sumbawan | Papuanesia |
| nna | Nyangumarta | nyan1301 | Pama-Nyungan | Western Pama-Nyungan | Australia |
| opm | Oksapmin | oksa1245 | Nuclear Trans New Guinea | Oksapmin | Papuanesia |
| pwn | Paiwan | paiw1248 | Austronesian | Paiwan | Papuanesia |
| pau | Palauan | pala1344 | Austronesian | Napuanesia |  |
| lkr | Päri | ari1256 | Nilotic | Romance | Africa |
| spa | Pasiego Spanish | stan1288 | Indo-European | Chinese | Eurasia |
| cmn | Pingding Mandarin | mand1415 | Sino-Tibetan | Peul-Serer | Eurasia |
| fuc | Pulaar | pula1263 | Atlantic-Congo | Puyuma | Africa |
| pyu | Puyuma | puyu1239 | Austronesian | Plavic | Papuanesia |
| rus | Russian | russ1263 | Indo-European | Billic | Papuanesia |
| bps | Sarangani Blaan | sara1326 | Austronesian | Atayalic | Papuanesia |
| trv | Seediq | Aaro1264 | Austronesian | Eurasia |  |
| sza | Semelai | seme1247 | Austroasiatic |  |  |


| ISO | Language | Glottocode | Family | WALS-Genus | Macro-Area |
| :--- | :--- | :--- | :--- | :--- | :--- |
| kca | Southwestern Khanty | khan1279 | Uralic | Ugric | Eurasia |
| erg | Sye | siee1239 | Austronesian | Oceanic | Papuanesia |
| tgl | Tagalog | taga1270 | Austronesian | Greater Central Philippine | Eurasia |
| tpj | Tapieté | tapi1253 | Tupian | Tupi-Guaraní | South America |
| ssf | Thao | thao1240 | Austronesian | Western Plains Austronesian | Papuanesia |
| suq | Tirmaga | suri1267 | Surmic | Surmic | Africa |
| bbc | Toba Batak | bata1289 | Austronesian | Northwest Sumatra-Barrier Islands | Papuanesia |
| dto | Tommo So Dogon | tomm1242 | Dogon | Dogon | Africa |
| tri | Trió | trio1238 | Cariban | Cariban | South America |
| tur | Turkish | nucl1301 | Turkic | Turkic | Eurasia |
| ulw | Ulwa | ulwa1239 | Misumalpan | Misumalpan | North America |
| uur | Ura | urav1235 | Austronesian | Nubian | Papuanesia |
| ghl | Uyncu | ghul1238 | Nubian | Wambayan | Africa |
| wmb | Wambaya | nucl1328 | Mirndi | Sahaptin | Australia |
| yak | Yakima Sahaptin | yaki1237 | Sahaptian | Wapei-Palei | North America |
| yev | Yeri | yapu1240 | Nuclear Torricelli | Zapotecan | Papuanesia |
| ctz | Zacatepec Chatino | zaca1242 | Otomanguean |  | North America |

Table 3: List of CoMDs in DoCoMD

| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| tbw1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| kjq1 | Inside a root, in final consonant position, $/ \mathrm{mm}^{\mathrm{S}} \mathrm{n} \mathrm{n}^{\mathrm{q}} \mathrm{b}$ d f g s dz s w j/ must occur. | "Final morphophemic consonants: N, M, M, W, Ẃ, B, Gw, Y, D, SD, Ş, Z, S" |
| aha1 | Inside a morpheme, /a/ cannot occur. | "There is every a priori possibility that Anufo (like its near relatives Ahanta, Akan, Anum, Chumburung, Gichode, Gonja, Krachi, and Nawuri) might have a [+ATR] allophone of $/ \mathrm{a} /$ in the vicinity of the $[+\mathrm{ATR}]$ vowels $/ \mathrm{i} /$, /e/, /o/, /u/." |
| aka1 | Inside a root, schwa cannot occur as the only vowel. | "/ə/, however, is subject to the restriction that it is never the only vowel in a root." |
| akk1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Akkadian allows only one labial per root" |
| akk2 | Inside a root, coronal consonants cannot occur with other labial consonants. | "Semitic roots avoid adjacent homorganic consonants [...]. One class of such consonants are coronal obstruents." |
| yor5 | Inside a morpheme, /I/, /v/ cannot occur. | "Allophonic [-ATR] dominance occurs in several Yoruba dialects. Like Standard Yoruba, these dialects, which include Akure [...], Ifaki [...], Ijesa [...], and Ekiti [...] [...] an allophonic process which realizes the [+ATR] vowels /i/ and /u/ as [-ATR] $[\mathrm{I}]$ and $[v]$ when these vowels precede syllables containing underlying [-ATR] vowels." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| rwm1 | Inside a morpheme, /e/ and /o/ cannot occur. | "There are a number of Bantu languages with such a seven-vowel system with ATR contrast only in the high vowels in the east of Congo, in Uganda and in Tanzania: Nande [...], Talinga/Bwisi [...], Amba [...], Gungu [...], Malila [...]. Such systems generally have the vowels [e] and [o] as allophones." |
| gwx1 | Inside a morpheme, /a/ cannot occur. | "There is every a priori possibility that Anufo (like its near relatives Ahanta, Akan, Anum, Chumburung, Gichode, Gonja, Krachi, and Nawuri) might have a [+ATR] allophone of $/ \mathrm{a} /$ in the vicinity of the $[+\mathrm{ATR}]$ vowels $/ \mathrm{i} / \mathrm{l} / \mathrm{e} / \mathrm{l} / \mathrm{o} / \mathrm{l} / \mathrm{u} / .{ }^{\prime \prime}$ |
| arb1 | Inside a root, identical consonants cannot occur as two non-rightmost consonants. | "In brief, Arabic allows roots of two, three, and four consonants, all of them subject to the Obligatory Contour Principle. Biconsonantal roots are realized on the surface with gemination of the second consonant as a direct consequence of the universal left-to-right association convention. Note also that the Obligatory Contour Principle excludes quadriliteral roots with adjacent identical autosegments, like hypothetical *ddrj or *drrj." |
| asm1 | Inside a morpheme, [+ATR] mid vowels do not occur. | "Thus, [e] and [o] occur only non-finally in the context of final [+ATR]." |
| avi1 | Inside a morpheme, /e/ and /o/ cannot occur. | "Avikam atr_harmony Ym" |
| avu1 | Inside a morpheme, /e/ and /o/ cannot occur. | "Avokaya atr_harmony Ym" |
| bdh1 | Inside a morpheme, /e/ and /o/ cannot occur. | "Baka atr_harmony Ym" |
| bct1 | Inside a morpheme, /e/ and /o/ cannot occur. | "Bendi atr_harmony Ym" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| bdi1 | Inside a morpheme, /e/ an /o/ cannot occur. | "Unlike all other vowel qualities, they virtually do not occur in monosyllabic words, the only exception encountered being the word cjooc 'five'. Apart from this exception, [e] and [o] occur only in syllables that are followed by a syllable containing either [i] or [u], i.e., a high [+ATR] voweI. |
| sro1 | Inside a root, /e/, /o/ cannot occur. | "Any phonological analysis of metaphony in CS must therefore exclude high-mid vowels in contrasting positions, like in the stressed position of a root" |
| yue1 | Inside a morpheme, a labial onset consonant cannot occur preceding a labial coda consonant in the same syllable. | "The last MSC holds between the nonadjacent onset and coda, which may not both be labial consonants." |
| cao1 | Inside a not occ | "The glottal fricative $/ \mathrm{h} /$ is restricted such that it never occurs at the end of a morpheme or stem." |
| cao2 | Inside a morpheme, sibilants cannot occur directly preceding a glottal stop or an approximant. |  and sibilant-glottal stop clusters (/sP/, / $\mathrm{S} \mathrm{R} / \mathrm{/} / \mathrm{s} \mathrm{S} / \mathrm{)}$ [...] are not attested intramorphemically in my database." |
| ncu1 | Inside a morpheme, /a/ cannot occur. | "There is every a priori possibility that Anufo (like its near relatives Ahanta, Akan, Anum, Chumburung, Gichode, Gonja, Krachi, and Nawuri) might have a [+ATR] allophone of $/ \mathrm{a} /$ in the vicinity of the $[+\mathrm{ATR}]$ vowels $/ \mathrm{i} /$, /e/, /o/, /u/." |
| njo1 | Inside a verb root, $\mathrm{HL}, \mathrm{HH}$ or MM must occur. | "HH, MM and HL in citation forms, the only possible and legitimate tonal patterns in underived verbs" |


| Index | CoMD | Quotation |
| :--- | :--- | :--- | :--- |
| yao1 | Inside a root, in initial vowel <br> position, long vowels must | "within a word any vowel found in morpheme initial position is long" |
|  |  |  |
| occur. |  |  | yao2 | Inside a morpheme, $/ \mathrm{m} /$ | "It should be mentioned that the syllabic nasal cannot occur morpheme internally." |
| :--- | :--- |
|  | cannot occur in medial posi- |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| tfn4 | Inside a root, at least two moras must occur. | "All stem syllables have to have an onset and the rhyme has to be at least bimoraic" |
| tfn4 | Inside a root, consonants cannot precede consonants in an onset. | "While stems themselves have no more than one onset consonant, complex onsets of stems syllables are possible as a result of affixation." |
| des1 | Inside a root, glottals must occur after the first vowel. | "Phonetic evidence regarding the occurrence of these laryngeal sounds show that the phonetic realization of [h] and [?] is better treated as single prosodic phenomena that occur after the first vowel within root morphemes." |
| des2 | Inside a root, /d/ cannot occur in medial position. | "[d]. It is in complementary distribution with [r]. [...] [d] occurs word-initially, and [r] occurs root internally" |
| did1 | Inside a morpheme, /a/ cannot occur. | "In one fairly common type of allophonic [+ATR] dominance, found in some 5 Ht and $4 \mathrm{Ht}(\mathrm{H})$ languages, the low [-ATR] vowel /a/ has a $[+\mathrm{ATR}]$ allophonic variant that occurs only when followed or preceded (depending on the language) by a [+ATR] vowel. Languages that display such a process include Akan (Kwa; [...]), Didinga (Surma; [...]), Kinande (Bantu; [...]), and Lugbara (Central Sudanic; [...])." |
| dji1 | Inside a morpheme, a consonant cannot occur directly preceding /r/. | "Djinang Intramorphemic consonant clusters second member r $\square$ " |
| dus1 | Inside a morpheme, /ss/ cannot occur. | "However, across the morpheme boundary, the alveolar fricative /s/ can also be geminated" |
| nld1 | Inside a morpheme, a cluster of voiced obstruents cannot occur. | "Dutch morphemes are subject to the constraint that voiced obstruent clusters only occur in complex words." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| mhr1 | Inside a morpheme, /æ/ cannot occur in initial syllables. | "The front low vowel [æ] is restricted to non-initial syllables, and surfaces only as a result of frontness harmony. There are no such vowels underlyingly." |
| etu1 | Inside a root, in final position, voiced plosives or nasals must occur. | "Final consonants are permitted in roots. However, they form a restricted set consisting of the three plosive $\mathrm{b}, \mathrm{d}$ and g , and the three nasals $\mathrm{m}, \mathrm{n}, \mathrm{g} . "$ |
| etu2 | Inside a morpheme, / / / cannot occur as the only vowel in a morpheme of the shape CV. | "In the final vowel position of -CV roots, any vowel may occur except the mid, back, unrounded vowel ə." |
| etu3 | Inside a root, as the final vowel in a CVCV root, /i/, $/ \varepsilon /$, and /a/ must occur. | "In the final position of $-\mathrm{CV}(\mathrm{C}) \mathrm{V}$ roots, the set of vowel is restricted to those that are front and unrounded. [...] Only two roots have been found with other than a front, unrounded vowel in a final position and in both cases the vowel is an a." |
| etu4 | Inside an affix, / / / or $/ \mathrm{u} /$ cannot occur in initial position. | here are no root initial vowels, but there are word initial vowels. In the case of uns and verbs these vowels are prefixes and are restricted to the set $\mathrm{i}, \varepsilon$, a, ว." |
| etu5 | Inside a morpheme, at least one syllable must occur. | "Morphemes are either monosyllabic or disyllabic." |
| etu6 | Inside a morpheme, more than two syllables cannot occur. | "Morphemes are either monosyllabic or disyllabic." |
| etu7 | Inside a root, $/ \mathrm{m} / \mathrm{h} / \mathrm{n} /, / \mathrm{y} /$ cannot occur in initial position directly preceding /u/ | "Also between the first consonant and the first vowel of roots, the nasals $m, \tilde{n}$ and $\eta$ do not occur before the back, rounded vowel u." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| etu8 | Inside a root, / $\partial$ / cannot occur directly preceding / $\mathrm{g} /$, /g/. | "In terms of the first vowel and the second consonant of noun and verb roots, the back, unrounded vowel ə never occurs before the velars $\mathrm{\eta}$ and g ." |
| etu9 | Inside a root, /y/ cannot occur adjacent to bilabial consonants. | "In terms of the first consonant and the first vowel of noun and verb roots, the consonants $\mathrm{p}, \mathrm{b}, \mathrm{kp}$, and m do not occur before the front, rounded vowel $\mathrm{u} .[\ldots]$ the front rounded vowel $u$ : namely, it does not occur adjacent to a bilabial." |
| etu10 | Inside an affix, high, low, or falling-to-low tones must occur. | "is found only with high, low and falling-to-low tones, because [...] only these three tones are found on such prefixes." |
| etu11 | Inside a morpheme, falling-to-mid or mid tones cannot occur in monosyllabic morphemes. | "The downstepped high or mid tone, and the falling-to-mid tone are not found on monosyllabic morphemes." |
| etu11 | Inside a morpheme, /e/ and /o/ cannot occur. | "First, a number of three-height five-vowel /izaju/ languages [...], including Pulaar [...], Ejagham, Gitonga, Tsonga, and Zulu, have surface 4Ht(M) seven-vowel [iecaəou] systems due to a $[+\mathrm{ATR}]$ spreading process that derives $[\mathrm{e}]$ and $[\mathrm{o}]$ from $/ \varepsilon /$ and $/ \mathrm{o} /$ in $[+$ ATR $]$ contexts." |
| yor2 | Inside a morpheme, /r/, /v/ cannot occur. | "Allophonic [-ATR] dominance occurs in several Yoruba dialects. Like Standard Yoruba, these dialects, which include Akure [...], Ifaki [...], Ijesa [...], and Ekiti [...] [...] an allophonic process which realizes the [+ATR] vowels /i/ and /u/ as [-ATR] $[\mathrm{r}]$ and $[v]$ when these vowels precede syllables containing underlying [-ATR] vowels." |
| eng1 | Inside a morpheme, a cluster of voiced obstruents cannot occur in final position. | "English morphemes, for instance, never end in a cluster of voiced obstruents" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| fmp1 | Inside a root, /w/ and / $\mathrm{P} /$ cannot occur in initial position. | "The following phonemes can occur in C1 position: /t/ /c/ /k/ /b/ /d/ /j/ /g/ /f/ /s/ /v/ /z/ /h/ /m/ /n/ /y/" |
| fmp2 | Inside a root, in final position, /b/, /d/, /g/, /h/, /R/, $/ \mathrm{m} /$ or $/ \mathrm{n} /$ must occur. | "Turning to C 2 position in the formula given in (1), it is noted that only the phonetic consonants $[\mathrm{p}],[\mathrm{t}],[\mathrm{k}],[\mathrm{m}],[\mathrm{n}],[\mathrm{h}]$ and $[\mathrm{p}] .[\ldots]$ The voiceless stops $[\mathrm{p}],[\mathrm{t}]$ and $[\mathrm{k}]$, however, are recognized as realizations of $/ \mathrm{b} / \mathrm{/} / \mathrm{d} /$ and $/ \mathrm{g} / .{ }^{\prime \prime}$ |
| fmp3 | Inside a root, high vowels cannot occur directly preceding /h/. | "The consonant $/ \mathrm{h} /$ permits only mid and low vowels in front of it, i.e. /e/, /a/, /o/ and $/ \mathrm{a} /$, but not $/ \mathrm{i} /, / \mathrm{u} /$ and $/ \mathrm{u} /$. Thus, a morpheme structure condition as in (58) is needed." |
| fmp4 | Inside a root, $/ \mathrm{b} /$ and $/ \mathrm{g} /$ must occur directly preceded by /o/ or /a/. | efore the [+grave] non-nasal stops /b/ and /g/ only /o/ and /a/ are found." |
| fmp5 | Inside a root, $/ \mathrm{m} /$ must occur directly preceded by /u/ or /a/. | "The two vowels / $\mathrm{u} /$ and /a/ occur before /m/" |
| fmp6 | Inside a root, /d/ must occur preceded by /a/. | "The only vowel that can occur before /d/ is /a/." |
| fmp7 | Inside a root, $/ \mathrm{n} /$ must occur preceded by $/ \mathrm{e} /$. | "The only vowel that occurs before /n/ is /e/." |
| fvr1 | Inside a morpheme, /e,o,a/ cannot occur. | "Fur Allophonic [+ATR] dominance $\boldsymbol{\nu}^{\prime \prime}$ |
| gaa1 | Inside a root, consonants cannot occur in final position. | "every primary root consists of an initial consonant and a vowel, a vowel cannot begin a root, a consonant cannot end it" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| gaa2 | Inside a root, vowels cannot occur in initial position. | "every primary root consists of an initial consonant and a vowel, a vowel cannot begin a root, a consonant cannot end it" |
| gaa3 | Inside a root, at least one syllable occurs. | "every primary root consists of an initial consonant and a vowel, a vowel cannot begin a root, a consonant cannot end it" |
| cab1 | Inside a root, /t/ cannot occur in initial position. | "In fact / $/$ / is absent in stem initial position in the native lexicon" |
| deu1 | Inside a root, at least two moras must occur. | "lexical root in German are minimally bimoraic" |
| deu2 | Inside a root, schwa cannot occur as the head of the initial syllable. | "there are no roots beginning with a syllable headed by a schwa" |
| acd1 | Inside a morpheme, /a/ cannot occur. | "There is every a priori possibility that Anufo (like its near relatives Ahanta, Akan, Anum, Chumburung, Gichode, Gonja, Krachi, and Nawuri) might have a [+ATR] allophone of /a/ in the vicinity of the [+ATR] vowels /i/, /e/, /o/, /u/." |
| toh1 | Inside a morpheme, /e/ and /o/ cannot occur. | "First, a number of three-height five-vowel /icaju/ languages [...], including Pulaar [...], Ejagham, Gitonga, Tsonga, and Zulu, have surface 4Ht(M) seven-vowel [ieعaวou] systems due to a $[+\mathrm{ATR}]$ spreading process that derives $[\mathrm{e}]$ and $[\mathrm{o}]$ from $/ \varepsilon /$ and $/ \mathrm{o} /$ in $[+A T R]$ contexts." |
| gis1 | Inside a morpheme, in the vowel position of a CV morpheme, /i u a/ must occur. | "Finally, nearly all vowels in morphemes of the form CV are cardinal vowels" |
| gjn1 | Inside a morpheme, /a/ cannot occur. | "There is every a priori possibility that Anufo (like its near relatives Ahanta, Akan, Anum, Chumburung, Gichode, Gonja, Krachi, and Nawuri) might have a [+ATR] allophone of $/ \mathrm{a} /$ in the vicinity of the $[+\mathrm{ATR}]$ vowels $/ \mathrm{i} /$, /e/, /o/, /u/." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| gni1 | Inside a root, $/ \mathrm{f} / \mathrm{h} / \mathrm{K} /$ cannot occur in initial position. | "The tap / dd/ and the laminal lateral /ly/ do not occur root initially." |
| gni2 | Inside an affix, / $\kappa, \mathrm{l}, \mathrm{d}, \mathrm{t}, \mathrm{n} /$ cannot occur in initial position. | "All consonants except /ly, rl, rd, th, and n/ are attested morpheme initially." |
| gni3 | Inside an affix, in final consonant position /n, n, m, r/ must occur. | "Just a handful of morphemes, all of which are pronominal prefixes, end in a consonant, which must be one of $/ \mathrm{n}, \mathrm{ny}, \mathrm{m}$, or $\mathrm{dd} / .{ }^{\prime \prime}$ |
| rub1 | Inside a morpheme, /e/ and /o/ cannot occur. | "There are a number of Bantu languages with such a seven-vowel system with ATR contrast only in the high vowels in the east of Congo, in Uganda and in Tanzania: Nande [...], Talinga/Bwisi [...], Amba [...], Gungu ([...]), Malila [...]. Such systems generally have the vowels [e] and [o] as allophones." |
| heb1 | Inside a root, identical consonants cannot occur as two non-rightmost consonants. | "No classical Semitic language contains stems [CiVCiX], where the left bracket marks the beginning of the stem and X is nonnull. [...] Finally, it is worth noting that there are no quadriliteral verb forms with doubling of any consonant except the final one" |
| qub1 | Inside a verb root, in final position, vowels must occur. | "Verbal roots must end in a vowel" |
| hun1 | Inside a morphem, short /o/, $/ \phi /$ cannot occur in final position. | "Short [-open1, +open2] vowels (/o, ö/) cannot occur in final position." |
| hun2 | Inside a root, short /i/, /y/, /u/, /e/, /a/ cannot occur in final position of monosyllabic roots. | "Domain final [-open2] vowels are (i) long in monosyllables" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| hun3 | Inside a root, at least two moras must occur. | "there is a constraint that applies to all dialects and requires that the minimal word/stem should be bimoraic" |
| hun4 | Inside a morpheme, long vowels cannot occur preceding CC consonant clusters. | "Any long vowel is possible before a cluster if there is a morpheme boundary after the vowel or between the consonants making up the cluster" |
| huz1 | Inside a root, at most two syllables can occur. |  |
| yor 4 | Inside a morpheme, /r/, /v/ cannot occur. | "Allophonic [-ATR] dominance occurs in several Yoruba dialects. Like Standard Yoruba, these dialects, which include Akure [...], Ifaki [...], Ijesa [...], and Ekiti [...] an allophonic process which realizes the [+ATR] vowels /i/ and /u/ as [-ATR] [r] and $[v]$ when these vowels precede syllables containing underlying [-ATR] vowels." |
| yor1 | Inside a morpheme, /I/, /v/ cannot occur. | "Allophonic [-ATR] dominance occurs in several Yoruba dialects. Like Standard Yoruba, these dialects, which include Akure [...], Ifaki [...], Ijesa [...], and Ekiti [...] [...] an allophonic process which realizes the $[+$ ATR] vowels $/ \mathrm{i} /$ and $/ \mathrm{u} /$ as $[-A T R]$ $[\mathrm{I}]$ and $[v]$ when these vowels precede syllables containing underlying [-ATR] vowels." |
| ind1 | Inside a root, palatal consonants cannot occur in final position. | "The condition states that morpheme-final consonants must have the feature specifications either [+Anterior, -Back] (i.e. dentals/alvolars and labials) or [-Anterior, +Back] (i.e. velars). Palatals (/š/, /č/, /y/), which have the feature specifications [-Anterior, -Back], do not satisfy the condition" |
| yor3 | Inside a morpheme, /I/, /v/ cannot occur. | "Allophonic [-ATR] dominance occurs in several Yoruba dialects. Like Standard Yoruba, these dialects, which include Akure [...], Ifaki [...], Ijesa [...], and Ekiti [...] [...] an allophonic process which realizes the [+ATR] vowels /i/ and /u/ as [-ATR] $[\mathrm{r}]$ and $[v]$ when these vowels precede syllables containing underlying [-ATR] vowels." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| ita1 | Inside a root, $[K]$ cannot occur in initial position. | "The palatal lateral $[\Lambda]$ is ruled out at the beginning of $\omega[+\mathrm{min}]$ " |
| mks1 | Inside a root, exactly two moras must occur. | "The canonical root in native vocabulary is bimoraic" |
| mks2 | Inside an affix, /i/ cannot occur. | "Only roots may be glottalized" |
| mks3 | Inside a root, $/ \mathrm{R} /$ must occur after the first vowel. | "contrastive glottalization is restricted to associate to the first vowel of the root" |
| ncal | Inside a morpheme, nasal vowels must occur as the rightmost vowel. | "Nasalisation in the Iyo language may be seen as a feature not of the segment but as a suprasegmental located on certain roots and suffixes. These roots and suffixes exist in the lexicon with a [+nasal] feature, realised phonetically on their final vowels." |
| jav1 | Inside a root, consonants cannot occur preceding a consonant with the same place and sonorant values, unless identical and being the first and second consonant. | "Labial consonants exclude other labial, as do coronal obstruents, coronal sonorants, palatal consonants and velar consonants. [...] identical consonants can in general cooccur in $\mathrm{C} 1 / \mathrm{C} 2{ }^{\prime \prime}$ |
| jav2 | Inside a root, alveolar consonants cannot occur as the first consonant preceding a palatal consonant as the second consonant. | "there is also one asymmetric constraint against coronal consonants in C1 combined with palatal consonants in C2" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| jum1 | Inside a morpheme, /e/ and /o/ cannot occur. | "In monosyllabic words, however, there are only eight vowel qualities, the mid [+ATR] vowels $[\mathrm{e}]$ and $[\mathrm{o}]$ not occurring in such words. Thus, $[\mathrm{e}]$ and [o] only occur in words in which an adjacent syllable contains /i/ or /u/, i.e. a high [+ATR] vowel, as illustrated in (5). Hence, [e] and [o] must be analysed as allophones of the [-ATR] phonemes $/ \varepsilon /$ and $/ \rho /$, respectively." |
| ksp1 | Inside a morpheme, $/ \varepsilon /$ cannot occur. | "A second form of allophonic [-ATR] dominance occurs in just a single survey l anguage, Kaba (Central Sudanic; [...]). Though it has three phonemic back round vowels $/ \mathrm{u} /, / \mathrm{\rho} / \mathrm{/o} / \mathrm{in}$ addition to the low central vowel $/ \mathrm{a} /$, and hence is considered a $4 \mathrm{Ht}(\mathrm{M})$ language for purposes of this study (see Note 2), Kaba lacks an [ATR] contrast among front vowels, in view of the fact that there is no phonemic nonhigh [-ATR] vowel $/ \varepsilon /$ in the language, the only front vowels being /i/ and $/ \mathrm{e} /$. Significantly, however, the vowel $[\varepsilon]$ does occur phonetically as an allophone of $/ \mathrm{e} /$ in syllables adjacent to those containing an underlying [-ATR] vowel / $\mathrm{a} /$ or / $\mathrm{o} / . \mathrm{}$. " |
| keo1 | Inside a morpheme, /e/ and /o/ cannot occur. | "There are two harmonic variants, [e], and [o]. When mid vowels occur with the $[+$ ATR $]$ vowel $[\mathrm{i}]$ or $[\mathrm{u}]$, then they become [+ATR] [e] or [o] as well. However, by themselves or with the [a], they remain as [-ATR] [ $\varepsilon$ ] or [ 0 ]." |
| kis1 | Inside a root, a coronal obstruent cannot occur preceding another coronal obstruent with the same manner of articulation but a different retroflexion. | "two coronal obstruents with the same manner of articulation [...] must agree with respect to retroflexion or non-retroflexion" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| xnb1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| pam1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| ckv1 | Inside a root, labial consonants cannot occur with other labial consonants. | "There are many possible combinations of consonants, but there is a tendency to avoid homorganic or back consonants, e.g. *pb-, *bp-, *bm-" |
| sst1 | Inside a root, a high tone cannot occur on the first syllable of a polysyllabic root. | "In isolation, no polysyllable occurs with a high tone in initial position |
| kzj1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| nnb1 | Inside a morpheme, /e/ and /o/ cannot occur. | "In some $4 \mathrm{Ht}(\mathrm{H})$ languages (Kinande, Talinga-Bwisi, Kirangi) [...], the fact that [ + ATR] mid vowels [e] and [o] do occur (allophonically) on the surface means that developing a detailed formal analysis that bans * $[\mathrm{e}], *[\mathrm{o}]$ as the output of coalescence but permits these vowels to arise via [+ATR] spreading is not necessarily a trivial exercise." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| nym1 | Inside a root, voiceless nasals, non-sibilant fricatives and voiceless plosives do not occur preceding voiceless nasals, non-sibilant fricatives or voiceless plosives. | "There are no roots of the structure $\mathrm{CV}(\mathrm{V}) \mathrm{C}(\mathrm{V})$ where both C's belong to the following set: $\mathrm{ptkfh} \mathrm{mh} \mathrm{nh} \mathrm{gh}{ }^{\prime \prime}$ |
| nym2 | Inside nominal affixes, mid vowels do not occur. | "In nominal and concording class prefixes, there is a system of five vowels i-1-a-v - u." |
| xom1 | Inside a morpheme, /e/, /o/, /a/ cannot occur. | "This process, henceforth [+ATR] harmony, spreads [+ATR] to [-high, -ATR] vowels $/ \varepsilon, \mathrm{a}, ~ \supset /$ causing the latter to surface as $[+\mathrm{ATR}]$ allophones $[\mathrm{e}, \partial, ~ o]$ respectively. These allophonic surface realizations only occur in domains preceding [+high, +ATR] vowels and do not occur in isolated monosyllabic roots." |
| kpy1 | Inside a morpheme, /w/ ca not occur in final position. | "The grammar presented above predicts that, if there were a morpheme ending in w , it would always be faithfully mapped to w . This would result in paradigms with no final consonant alternations [...] However, no such morphemes exist: whenever a morpheme-final segment surfaces as w when not followed by a vowel, that segment surfaces as v when followed by a vowel." |
| kye1 | Inside a morpheme, /ą/ cannot occur. | "There is every a priori possibility that Anufo (like its near relatives Ahanta, Akan, Anum, Chumburung, Gichode, Gonja, Krachi, and Nawuri) might have a [+ATR] allophone of $/ \mathrm{a} /$ in the vicinity of the $[+\mathrm{ATR}]$ vowels $/ \mathrm{i} /$, /e/, /o/, /u/." |
| bfa1 | Inside a root, doubly articulated consonants and voiced obstruents cannot occur in final position. | "root final consonants. In general, neither voiced obstruents nor, in the terms of the feature geometry [...], doubly linked consonants can occur in this position." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| lkt1 | Inside a root, closed syllables cannot occur. | "codas are generally banned in roots, but they may surface in affixes, function words, and reduplicants" |
| las1 | Inside a morpheme, /a/ cannot occur. | "The low vowel /a/ does not have an underlying [+ATR] counterpart. [...] all the vowels contrast minimally in terms of [ATR] in CV words. Such a contrast does not exist in the vowel/a/. However, there is an [+ATR] low vowel [ $\Lambda]$ found only in the environment of other [+ATR] vowels." |
| lag1 | Inside a morpheme, /e/ and /o/ cannot occur. | "In some $4 \mathrm{Ht}(\mathrm{H})$ languages (Kinande, Talinga-Bwisi, Kirangi) [...], the fact that [+ATR] mid vowels [e] and [o] do occur (allophonically) on the surface means that developing a detailed formal analysis that bans *[e], *[o] as the output of coalescence but permits these vowelsto arise via $[+\mathrm{ATR}]$ spreading is not necessarily a trivial exercise." |
| laj1 | Inside a root, consonant clusters must occur in initial position. | "CG sequences occur only root initially." |
| laj2 | Inside a root, /w/ must occur as part of the initial onset. | "[w] occurs in morpheme initial position and in clusters following any consonant morphophoneme except /b/ and the glide /y/ in morpheme initial position. [...] [w] never occurs in other positions with roots." |
| laj3 | Inside a root, fricatives, geminates and /r/ cannot occur in initial position. | "Second, a number of sounds are only found in C: 1) geminates 2) fricatives, 3) [r]" |
| laj4 | Inside an affix, /o/, /o/ and $/ v /$ cannot occur. | "All vowels occur in suffixes, but no prefix contains $/ \mathrm{\rho} / \mathrm{/} / \mathrm{v} / \mathrm{L} / \mathrm{u} / . \mathrm{"}$ |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| laj5 | Inside a root, high vowels canot occur in initial position. | "Only non-high vowel can occur root initially" |
| les1 | Inside a morpheme, /e/, /o/ cannot occur | "Lese-Mvuba Allophonic [+ATR] dominance $\boldsymbol{\checkmark}$ " |
| bod1 | Inside a morpheme, /i/ cannot occur. | "this vowel is not underlying. [...] this vowel does not trigger raising, but it can also result from raising" |
| kmk1 | Inside a root, labial consonants cannot occur as the initial consonant preceding a labial consonant. | "Acehnese can be analyzed, like Limos Kalinga, as having *[stemLAB V0 LAB" |
| 1 lq 1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| dni1 | Inside a morpheme, $/ \mathrm{y} /$ and $/ \mathrm{R} /$ cannot occur in initial position. | "Morpheme initially any single consonant except glottal stop or $/ \mathrm{g} /$, the velar nasal, may occur." |
| dni2 | Inside a morpheme, $/ \mathrm{k}^{\mathrm{w}} /$, /g/, /h/, /w/ and /j/ cannot occur in medial position. | "Morpheme medially any single consonant except $/ \mathrm{k}^{\mathrm{w}} / \mathrm{h} / \mathrm{h} / \mathrm{/} / \mathrm{g} / \mathrm{/} / \mathrm{w} /$ and $/ \mathrm{j} /[\ldots]$ may occur" |
| dni3 | Inside a morpheme, geminate $/ \mathrm{k}^{\mathrm{w}}$ / cannot occur. | "The following types of clusters of two consonants occur morpheme medially: (1) All stops except /kw/ occur in geminate cluster" |


| Index | CoMD | Quotation |
| :--- | :--- | :--- |
| dni4 | Inside a morpheme, plosives <br> cannot occur directly preced- | "The cluster /ks/ is the only recorded example of stop followed by voiceless continuant |
| ing /h/. |  |  |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| mgq1 | Inside a morpheme, /e/ and /o/ cannot occur. | "There are a number of Bantu languages with such a seven-vowel system with ATR contrast only in the high vowels in the east of Congo, in Uganda and in Tanzania: Nande [...], Talinga/Bwisi [...], Amba [...], Gungu [...], Malila [...]. Such systems generally have the vowels $[\mathrm{e}]$ and $[\mathrm{o}]$ as allophones." |
| mdk1 | Inside a morpheme, /e/, /o/ cannot occur | "Mangbutu atr_harmony Ym" |
| arn1 | In a root, at most one consonant cluster can occur. |  |
| vma1 | Inside a morpheme, a consonant cannot occur directly preceding /t n n/. | "The second member of such a cluster is drawn from the set of peripheral consonants plus the palatal glide $/ \mathrm{y} /$; that is, a subset of the consonants permitted in initial position." |
| mdm1 | Inside a morpheme, /e/ and /o/ cannot occur. | "A second common form of allophonic [+ATR] dominance involves the phonetic realization, in a number of $4 \mathrm{Ht}(\mathrm{H})$ systems, of underlying [-ATR] mid vowels $/ \varepsilon /$ and $/ \mathrm{o} /$ as $[+$ ATR $]$ allophones $[\mathrm{e}]$ and $[\mathrm{o}]$ respectively when these underlying vowels occur in a syllable adjacent to underlying $[+$ ATR $]$ vowels $/ \mathrm{i} / \mathrm{/} / \mathrm{u} /$. Such a process is attested for example in LuBwisi (Bantu; [...]), Kinande (Bantu; [...]), and Mayogo (Ubangi; [...]). |
| tay 1 | Inside a root, labial consonants cannot occur as the initial consonant preceding a labial consonant. | "And, labials later in the stem still have no effect (e.g., Mayrinax Atayal q-um-ami). We invoke the constraint mentioned in the introduction, *[stemLAB V0 LAB, which forbids a stem-initial labial when the next consonant is labial." |
| jkr1 | Inside a morpheme, diphthongs cannot occur. | the derived syllable structure, XV can be different from $\mathrm{V}^{\text {" }}$ |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| jkr2 | Inside a morpheme, superheavy syllables cannot occur. | "a derived syllable structure exists due to the affixation and compounding of morphemes that results in a sequence which is longer than the sequences given above, but still not clearly two syllables. The result is a derived (C)V-V(C) structure which can be realised as a single syllable." |
| mnb1 | Inside a root, labial obstruents cannot occur as the first or second consonant with the other consonant being a labial obstruent with the opposite voicing value. | "Initial plosives do not co-occur with contra-voiced homorganic plosives [...] For the labials the rule can therefore be extended to exclude all cooccurrences of contra-voiced labial obstruents" |
| mnb2 | Inside a root, plosives cannot occur as the first or second consonant with the other consonant being a homoorganic plosive with the opposite voicing value. | "Initial plosives do not co-occur with contra-voiced homorganic plosives" |
| mnb3 | Inside a root, prenasalized plosives cannot occur as the first or second consonant with the other consonant being a homoorganic obstruent with the opposite voicing value. | "Prenasalized plosives do not co-occur with contra-voiced obstruents." |


| Index | CoMD | Quotation |
| :--- | :--- | :--- |
| mnb4 | Inside a root, obstruents and | "Obstruents and prenasalized consonants do not co-occur with homorganic nasals." |
|  | prenasalized plosives cannot <br> occur as the first or second |  |
|  | consonant with the other |  |
| consonant being a homoor- |  |  |
| ganic nasal. |  |  |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| sur3 | Inside a morpheme, glides, glottalized consonants, voiced consonants, affricates and labial fricatives cannot occur in final position. | "There are no glides, no voiced or glottalized consonants, no affricates, and no labial fricatives in morpheme final position." |
| sur4 | Inside a morpheme, a high vowel cannot occur preceding a mid vowel. | "there may not be a sequence of a mid vowel followed by a high vowel." |
| acx1 | Inside a root, a consonant cannot occur preceding a homorganic consonant except if identical. | "In MA, as in other Arabic dialects, phonotactic constraints among consonants are enacted primarily as the morpholexical level, where triconsonantal and quadriconsonantal roots are restricted with few exceptions to sets of non-homoorganic consonants, even in non-adjacent radical positions (e.g. the first and third radical of a triradical)." |
| acx2 | Inside an affix, back or emphatic consonants cannot occur. | "At the same time, consonants acting as non-radicals, that is occuring in derivational and inflectional affixes and in clitics -/b, h, k, l, m, n, t, y, $\mathrm{l} /$ - exclude the back (except for the low back) consonants / $\mathrm{q}, \mathrm{x}, \mathrm{y}, \mathrm{\hbar}, \mathrm{f} /$ and emphatic consonants / O , t, s/" |
| mxh1 | Inside a morpheme, /e/, /o/ cannot occur | "Lese-Mvuba Allophonic [+ATR] dominance $\boldsymbol{\vee}$ " |
| jup1 | Inside a morpheme, $/ \mathrm{j} /, / \mathrm{g} /$, and $/ \mathrm{c} /$ must occur in final consonant position | "Three of Hup's consonants can only appear in morpheme-final position: /j/, /g/, and /ç/." |
| naw1 | Inside a morpheme, /ą/ cannot occur. | "There is every a priori possibility that Anufo (like its near relatives Ahanta, Akan, Anum, Chumburung, Gichode, Gonja, Krachi, and Nawuri) might have a [+ATR] allophone of $/ \mathrm{a} /$ in the vicinity of the $[+\mathrm{ATR}]$ vowels $/ \mathrm{i} /$, /e/, / $\mathrm{o} / \mathrm{/} / \mathrm{u} / .{ }^{\prime \prime}$ |


| Index | CoMD | Quotation |
| :--- | :--- | :--- |
| niy1 | Inside a morpheme, labial af- <br> fricates must occur directly <br> preceding rounded vowels. | in monomorphemic foms." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| ace1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| nso1 | Inside a morpheme $/ v /$ cannot occur. | "The languages would appear to have an extra (allophonic) height series, which comes from raising non-advanced high vowels to higher vowel that are distinct from the high advanced vowels." |
| nna1 | Inside a noun root, short vowels cannot occur in monosyllabic roots. | "A very small number of roots is monosyllabic. Such roots, if nouns, contain long vowels but not short vowels [...] or, if verbs, contain short vowels but not long vowels [...]" |
| nna2 | Inside a verb root, long vowels cannot occur in monosyllabic roots. | "A very small number of roots is monosyllabic. Such roots, if nouns, contain long vowels but not short vowels [...] or, if verbs, contain short vowels but not long vowels [...]" |
| nna3 | Inside a morpheme, as the first consonant of a consonant cluster, alveolar sonorants, palatal sonorants or retroflex consonants must occur. | "Some restrictions exist as to the privileges of occurrence of clusters within the word across morpheme boundaries. However, the following clusters, absent from category (3) below, are recorded in this environment: /pk pm ṭty ṭm tyk ṇy ly lm lny lw lym lyw nyy nw/" |
| nna4 | Inside a morpheme, as the second consonant of a consonant cluster, velar consonants or a labial consonants must occur. | "Some restrictions exist as to the privileges of occurrence of clusters within the word across morpheme boundaries. However, the following clusters, absent from category (3) below, are recorded in this environment: /pk pm ṭty ṭm tyk ṇy ly lm lny lw lym lyw nyy nw/" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| nzb1 | Inside a morpheme, /e/ and /o/ cannot occur. | "This is confirmed by the fact that raising of low-mid to high vowels is allophonic" |
| opm1 | Inside a morpheme, /lt/ cannot occur. | "A number of clusters which appear to be illicit within a single morpheme are permitted in a single phonological word which consists of more than one morphological word, e.g. /lt/" |
| $\operatorname{lgn} 1$ | Inside a morpheme, /e/, /o/, /a/ cannot occur. | "Phonemically, Tapo has three front, three back and one mid contrastive vowels. These are the contrastive $[+$ High, + ATR $] / \mathrm{i}, \mathrm{u} /,[+\mathrm{High},-\mathrm{ATR}] / \mathrm{r}, v /$, the $[+\mathrm{Mid}$, -ATR] $/ \varepsilon, \rho /$ and the open $[-A T R] / a /$. Phonetically, the language has nine vowels with $[+\mathrm{Mid},+$ ATR $][\mathrm{e}, \mathrm{o}]$ and $[+$ Low, + ATR $][\theta]$ that occur as a result of + ATR spreading to the $/ \varepsilon, \rho, \mathrm{a} /{ }^{\prime}$ |
| pwn1 | Inside a root, labial consonants cannot occur as the initial consonant preceding a labial consonant. | "The Kulalao Paiwan pattern can be analyzed with the ranking *[stemLAB V0 LAB » IDENT(place)/affix." |
| pau1 | Inside a root, labial consonants cannot occur with other labial consonants. | "I propose that there exists a constraint against more than one labial consonant within the root domain" |
| lkr1 | Inside an affix, mid vowels do not occur. | "Mid vowels do not occur in prefixes." |
| spa1 | Inside a morpheme, $/ \varepsilon /$ cannot occur" | "the missing [ $\varepsilon$ ] must be derived at an intermediate level" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| cmn1 | Inside a root, liquids cannot occur as the second consonant of an onset cluster. | "To begin with, the syllable structure of Chinese languages is straightforwardly (C) (G)V(C), where 'G' stands for a glide (i.e., j or w). Thus, it is surprising that infixation should create onset clusters which are otherwise not attested elsewhere in the language. On top of that, a retroflex lateral is not commonly found in descriptions of Mandarin phonetic inventory [...]. The appearance of a retroflex lateral only in forms with infixation also demands an explanation." |
| cmn2 | Inside a root, retroflex laterals cannot occur. | "To begin with, the syllable structure of Chinese languages is straightforwardly (C) (G)V(C), where 'G' stands for a glide (i.e., j or w). Thus, it is surprising that infixation should create onset clusters which are otherwise not attested elsewhere in the language. On top of that, a retroflex lateral is not commonly found in descriptions of Mandarin phonetic inventory [...]. The appearance of a retroflex lateral only in forms with infixation also demands an explanation." |
| fuc1 | Inside a morpheme, /e/ and /o/ cannot occur. | "First, a number of three-height five-vowel /icaju/ languages [...], including Pulaar [...], Ejagham, Gitonga, Tsonga, and Zulu, have surface 4Ht(M) seven-vowel [iecaəou] systems due to a $[+\mathrm{ATR}]$ spreading process that derives $[\mathrm{e}]$ and $[\mathrm{o}]$ from $/ \varepsilon /$ and $/ \mathrm{\rho} /$ in $[+$ ATR $]$ contexts." |
| puu1 | Inside a morpheme, /e/ and /o/ cannot occur. | "This language has five vowels $[\mathrm{i}, \varepsilon, \mathrm{u}, \rho, \mathrm{a}]$. Mid vowels are realized as $[\mathrm{e}, \mathrm{o}]$ before [i] and [u]" |
| pyu1 | Inside a root, labial consonants cannot occur with other labial consonants. | "The bilabial nasal sound $/ \mathrm{m} /$ becomes $/ \mathrm{n} /$ when the infix is affixed to a bilabial stop. [...] Unlike the assimilation described earlier, this rule is obligatory." |
| rus1 | Inside a morpheme, in vowel sequences, /iu/ or /au/must occur. | "only two vowel sequences are admitted within a Russian morpheme; they are $\left.\right\|^{*} \mathrm{i}^{*} \mathrm{u} \mid$ or $\left\|* a^{*} u\right\| "$ |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| bps1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| mbs1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| $\operatorname{trv} 1$ | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| sza1 | Inside a morpheme, in the coda of the penultimate syllable, /m, n, y, l, r/ must occur. | "In simple monomorphemic disyllabic words the penult is either open with an empty coda, or in the case of a heavy penult, the coda can only be filled by a sonorant phoneme: /m, n, g, l, r/." |
| sza2 | Inside a root, $/ \mathrm{u} / \mathrm{l} / \mathrm{\rho} /$ cannot occur as the final vowel in monosyllabic roots. | "All oral vowels except the central vowels $/ \mathbf{u} /$ and $/ \partial /$ occur in open final syllables in monosyllabic roots" |
| sot1 | Inside a morpheme, /I/ and /v/ cannot occur. | "The languages would appear to have an extra (allophonic) height series, which comes from raising non-advanced high vowels to higher vowel that are distinct from the high advanced vowels." |
| tsn1 | Inside a morpheme, /I $/$ and $/ v /$ cannot occur. | "The languages would appear to have an extra (allophonic) height series, which comes from raising non-advanced high vowels to higher vowel that are distinct from the high advanced vowels." |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| kca1 | Inside a morpheme, /y/, $/ \mathrm{u} /$, /æ/ cannot occur in initial syllables. | "Vowels [ī], [ä], and [ü] do not appear contrastively, and never appear in initial syllables, but only as results of vowel harmony, to be discussed next." |
| erg1 | Inside a root, at least one syllable occurs. | "Morphologically simple free forms can have the minimal shape of just a single syllable." |
| erg2 | Inside a root, /w/ cannot directly precede $/ \mathrm{u} /$. | "/wu/ is prohibited intramorphemically" |
| erg3 | Inside a verb root, a velar consonant, /h/ or /n/ cannot occur in initial position. | "there are no roots beginning velar consonants, the fricative /h/ , or the nasal /n-/." |
| $\operatorname{erg} 4$ | Inside a verb root, consonant clusters cannot occur in initial position. | "verbal roots never begin with consonant clusters" |
| erg5 | Inside a root, glides cannot occur in final position. | "Roots of all kinds can end in any vowel, or in any single consonant except the glides /y/ and /w/." |
| erg6 | Inside a root, /nl/ cannot occur in initial position. | "with the further restriction that there can be no root-initial sequence of $/ *_{n s}$-/, $/{ }^{*} n h-/$ and / ${ }^{n}$ nl-/." |
| erg7 | Inside a root, /h/ cannot occur directly preceded by any consonant except / $\mathrm{y} /$ | "/у/ /h/ +" |
| erg8 | Inside a root, /p/ cannot be directly preceded by $/ \mathrm{p} /$, /k/, /n/. | "/p/ /p/ $\square / \mathrm{k} / \square / \mathrm{y} / \square{ }^{\text {a }}$ |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| tgl1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| tlj1 | Inside a morpheme, /e/ and /o/ cannot occur. | "In some $4 \mathrm{Ht}(\mathrm{H})$ languages (Kinande, Talinga-Bwisi, Kirangi) [...], the fact that [+ATR] mid vowels [e] and [o] do occur (allophonically) on the surface means that developing a detailed formal analysis that bans $*[\mathrm{e}], *[\mathrm{o}]$ as the output of coalescence but permits these vowelsto arise via [+ATR] spreading is not necessarily a trivial exercise." |
| tpj1 | Inside a root, $/ \mathrm{s} /$ and $/ \mathrm{J} /$ do not occur together. | "Likely, there is a tendency for the alveolar fricative $[\mathrm{s}]$ and the palatal fricative $[\mathrm{J}]$ to co-occur more often with voiceless stops avoiding the combination with each other." |
| ssf1 | Inside a root, labial consonants cannot occur with other labial consonants. | "/b/ do es not occur in the same morpheme with /p/ (hence **bVp, **pVb)." |
| tih1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |
| suq1 | Inside an affix, in initial consonant position, coronal consonants must occur. | "The consonant-initial suffixes always begin with a coronal consonant." |
| bbc1 | Inside a root, labial consonants cannot occur with other labial consonants. | "Monomorphemic stems largely obey the OCP: in Tagalog it is very rare to find two labial consonants within an unreduplicated root, as presumably in the other languages here too" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| dto1 | Inside a root, geminate [ $1:]$ and [y:] cannot occur. | "The distribution of sonorant geminates is further restricted in stems to nasals and one instance of geminate /ww/. Geminates [11] and [yy] can arise from the morphophonology," |
| dto2 | Inside a root, $/ \mathrm{p} /$, /t/, /k/ cannot occur in medial position. | "Voiceless obstruents, especially stops, cannot occur stem-internally," |
| dto3 | Inside a root, /j/ cannot occur intervocalically. | "This section discusses constraints on individual consonants stem-internally rather than word-internally, since the category of words contains crypto-compounds and other derived forms that may have looser phonotactics than stems. [...] the alveolopalatal affricate $/ \mathrm{j}$ / cannot occur intervocalically" |
| dto4 | Inside a root, nasal vowels must occur in monosyllabic roots. | "In native Tommo So stems, nasal vowels only occur in monosyllabic stems; they must be the only vowel." |
| dto5 | Inside a root, at least one high tone must occur. | "so every stem, be it nominal, verbal, or adjectival, must have at least one H tone (and at most one H tone stretch)." |
| dto6 | Inside a root, at most one high tone stretch cannot occur. | "so every stem, be it nominal, verbal, or adjectival, must have at least one H tone (and at most one H tone stretch)." |
| dto7 | Inside a noun root, a high tone must occur on the final TBU. | "In this section, I group together all non-verbal stems, since their tonal behavior is the same. In native stems, this H tone stretch must be right-aligned. That is, the H tone stretch always begins at the right edge." |
| tri1 | Inside a root, /t/ cannot occur directly preceding /e/. | "The sequence /te/ which occured in the text sample, but did not in the word list always includes a morpheme break" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| tso1 | Inside a morpheme, /o/ cannot occur. | "First, a number of three-height five-vowel /izasu/ languages [...], including Pulaar [...], Ejagham, Gitonga, Tsonga, and Zulu, have surface $4 \mathrm{Ht}(\mathrm{M})$ seven-vowel [iecaəou] systems due to a $[+\mathrm{ATR}]$ spreading process that derives $[\mathrm{e}]$ and $[\mathrm{o}]$ from $/ \varepsilon /$ and $/ \mathrm{o} /$ in [ + ATR] contexts." |
| tur1 | Inside a root, an unrounded high vowel cannot occur directly preceded by a sequence of a low vowel and a labial consonant. | "a high, short harmonic vowel is rounded in the second syllable of a disyllabic word whose first vowel is $/ \mathrm{a} /$, and whose medial consonant cluster contains a labial /p,b,m,v/, and it is then de-harmonized." |
| tyv1 | Inside a root, an alveolar sibilant fricative cannot ocur across a vowel next to a postalveolar sibilant fricative. | "if a Tuvan root contains two sibilant fricatives that are separated from each other by a vowel, we find that the sibilants are always either both alveolar or both palatoalveolar" |
| ulw1 | Inside a morpheme, geminates cannot occur. | "There are no tautomorphemic geminates on the surface, but there are many examples of geminates at morpheme boundaries" |
| uur1 | Inside a root, at least one syllable occurs. | "Morphologically simple free forms in Ura can have the minimal shape of just a single syllable" |
| uur2 | Inside a noun root, $/ \mathrm{y} /$ and /h/ cannot occur in initial position. | "All consonants except /h/ are attested initially [...] While initial $/ \mathrm{y} /$ is attested on a number of words belonging to closed word classes, no noun is attested with this initial segment" |
| uur3 | Inside a verb root, in initial consonant position, $/ \mathrm{t} / \mathrm{/} / \mathrm{s} /$, $/ \mathrm{v} / \mathrm{L} / \mathrm{w} /$ or $/ \mathrm{j} /$ must occur. | "The only consonants that are commonly found verb-initially are the alveolar consonants /t-/ and /s-/, while a handful of verb roots begin with the glides /w-/ and /y-//, and the fricative /v-/" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| uur4 | Inside a root, non-sibilant fricatives cannot occur in final position. | "there is therefore a systematic exclusion against the fricatives $/ \mathrm{v} /, / \mathrm{f} / \mathrm{h} / \mathrm{h} /$ and $/ \mathrm{y} /$ in root-final position" |
| ghl1 | Inside a morpheme, /e/ and /o/ cannot occur. | ny Ym" |
| wmb1 | Inside a noun root, in final consonant position /c n k/ must occur. | "There are a few consonant-final nominal roots (involving /j/, /g/ and /ny/)" |
| wmb2 | Inside a verb root, in fina consonant position / t c k n j y f l/ must occur | "many of the verbal stems are consonant-final ending in either $/ \mathrm{n} /, / \mathrm{rd} /, / \mathrm{rl} /, / \mathrm{rr} /$, /j/, /ny/, /g/, or /ng/." |
| $\operatorname{lgg} 1$ | Inside a morpheme, /e/ and /o/ cannot occur. | "The phonemes $/ \varepsilon /$ and $/ \rho /$ are phonetically $[\mathrm{e}]$ and $[\mathrm{o}]$, respectively, when they cooccur with /i/ or $/ \mathrm{u} /$ within a word." |
| yev1 | Inside a root, nasal consonants cannot occur in the coda of a non-final syllable. | "For instance, very few nouns allow word-internal nasal codas and most of these involve reduplications [...]. Furthermore, although verbs permit nasal codas after the first syllable, these are morphemic, and involve either the third person singular masculine object morpheme or the imperfective morpheme" |
| yev2 | Inside a root, superheavy syllables cannot occur. | "Finally, coda consonants can occur with VV sequences, though this is uncommon. The vast majority of examples where this occurs involve ideophones (e.g. drualdrual /drualdrual/ 'slip', niryiar /niargiar/ 'squeak') or verb roots which have first syllable coda consonants and which select for the irrealis vowel /ie/" |
| ctz1 | Inside a root, /e/ cannot occur in non-final syllables. | "/e/ never occurs underlyingly in non-final syllables of roots" |


| Index | CoMD | Quotation |
| :---: | :---: | :---: |
| zne1 | Inside a morpheme, /e/ and /o/ cannot occur. | "First, a number of three-height five-vowel /izasu/ languages [...], including Pulaar [...], Ejagham, Gitonga, Tsonga, and Zulu, have surface 4Ht(M) seven-vowel [iecaəou] systems due to a $[+\mathrm{ATR}]$ spreading process that derives $[\mathrm{e}]$ and $[\mathrm{o}]$ from $/ \varepsilon /$ and $/ \mathrm{\rho} /$ in $[+$ ATR $]$ contexts." |
| zul1 | Inside a morpheme, /e/ and /o/ cannot occur. | "First, a number of three-height five-vowel /izaju/ languages [...], including Pulaar [...], Ejagham, Gitonga, Tsonga, and Zulu, have surface 4Ht(M) seven-vowel [iecaəou] systems due to a $[+\mathrm{ATR}]$ spreading process that derives $[\mathrm{e}]$ and $[\mathrm{o}]$ from $/ \varepsilon /$ and $/ \mathrm{o} /$ in [ +ATR] contexts." |

Table 4: Constraint Properties in DoCoMD

| Index | Substance | Type | Domain |
| :---: | :---: | :---: | :---: |
| tbw1 | Consonant | Sequence | root |
| kjq1 | Consonant | Positional | root |
| aha1 | Vowel | Inventory | morpheme |
| aka1 | Vowel | Inventory | root |
| akk1 | Consonant | Sequence | root |
| akk2 | Consonant | Sequence | root |
| yor5 | Vowel | Inventory | morpheme |
| rwm1 | Vowel | Inventory | morpheme |
| gwx1 | Vowel | Inventory | morpheme |
| arb1 | Consonant | Sequence | root |
| asm1 | Vowel | Inventory | morpheme |
| avi1 | Vowel | Inventory | morpheme |
| avu1 | Vowel | Inventory | morpheme |
| bdh1 | Vowel | Inventory | morpheme |
| bct1 | Vowel | Inventory | morpheme |
| bdi1 | Vowel | Inventory | morpheme |
| sro1 | Vowel | Inventory | root |
| yue1 | Consonant | Sequence | morpheme |
| cao1 | Consonant | Positional | morpheme |
| cao2 | Consonant | Sequence | morpheme |
| ncu1 | Vowel | Inventory | morpheme |
| njo1 | Tone | Sequence | root |
| yao1 | Vowel | Positional | root |
| yao2 | Consonant | Positional | root |
| yao3 | Vowel | Positional | root |
| com1 | Consonant+Vowel | Sequence | morpheme |
| tfn1 | Consonant | Inventory | affix |
| tfn2 | Vowel | Inventory | affix |
| tfn3 | Tone | Inventory | root proper |
| tfn4 | Mora | Minimality/Maximality | root proper |
| tfn4 | Consonant | Positional | root proper |
| des1 | Consonant+Vowel | Positional | root proper |
| des2 | Vowel | Positional | root proper |
| did1 | Vowel | Inventory | morpheme |
| dji1 | Consonant | Sequence | morpheme |
| dus1 | Consonant | Inventory | morpheme |


| Index | Substance | Type | Domain |
| :---: | :---: | :---: | :---: |
| nld1 | Consonant | Sequence | morpheme |
| mhr1 | Vowel | Positional | morpheme |
| etu1 | Consonant | Positional | root proper |
| etu2 | Vowel | Positional | morpheme |
| etu3 | Vowel | Positional | root |
| etu4 | Vowel | Positional | affix |
| etu5 | Syllable | Minimality/Maximality | morpheme |
| etu6 | Syllable | Minimality/Maximality | morpheme |
| etu7 | Consonant+Vowel | Positional | root |
| etu8 | Consonant+Vowel | Sequence | root |
| etu9 | Consonant+Vowel | Sequence | root |
| etu10 | Tone | Inventory | affix |
| etu11 | Tone | Inventory | morpheme |
| etu11 | Vowel | Inventory | morpheme |
| yor2 | Vowel | Inventory | morpheme |
| eng1 | Consonant | Positional | morpheme |
| fmp1 | Consonant | Positional | root proper |
| fmp2 | Consonant | Positional | root proper |
| fmp3 | Consonant+Vowel | Sequence | root |
| fmp4 | Consonant+Vowel | Sequence | root |
| fmp5 | Consonant+Vowel | Sequence | root |
| fmp6 | Consonant+Vowel | Sequence | root |
| fmp 7 | Consonant+Vowel | Sequence | root |
| fvr1 | Vowel | Inventory | morpheme |
| gaa1 | Consonant | Positional | root |
| gaa2 | Vowel | Positional | root |
| gaa3 | Syllable | Minimality/Maximality | root |
| cab1 | Consonant | Positional | root proper |
| deu1 | Mora | Minimality/Maximality | root proper |
| deu2 | Syllable | Positional | root proper |
| acd1 | Vowel | Inventory | morpheme |
| toh1 | Vowel | Inventory | morpheme |
| gis1 | Vowel | Inventory | morpheme |
| g.jn1 | Vowel | Inventory | morpheme |
| gni1 | Consonant | Positional | root proper |
| gni2 | Consonant | Positional | affix |
| gni3 | Consonant | Positional | affix |
| rub1 | Vowel | Inventory | morpheme |


| Index | Substance | Type | Domain |
| :---: | :---: | :---: | :---: |
| heb1 | Consonant | Sequence | root |
| qub1 | Vowel | Positional | V root proper |
| hun1 | Vowel | Positional | morpheme |
| hun2 | Vowel | Positional | morpheme |
| hun3 | Mora | Minimality/Maximality | root proper |
| hun4 | Consonant+Vowel | Sequence | morpheme |
| huz1 | Syllable | Minimality/Maximality | root |
| yor4 | Vowel | Inventory | morpheme |
| yor1 | Vowel | Inventory | morpheme |
| ind1 | Consonant | Positional | morpheme |
| yor3 | Vowel | Inventory | morpheme |
| ita1 | Consonant | Positional | root |
| mks1 | Mora | Minimality/Maximality | root proper |
| mks2 | Consonant | Inventory | affix |
| mks3 | Consonant+Vowel | Positional | root |
| nca1 | Vowel | Positional | morpheme |
| jav1 | Consonant | Sequence | root |
| jav2 | Consonant | Sequence | root |
| jum1 | Vowel | Inventory | morpheme |
| ksp1 | Vowel | Inventory | morpheme |
| keo1 | Vowel | Inventory | morpheme |
| kis1 | Consonant | Sequence | root |
| xnb1 | Consonant | Sequence | root |
| pam1 | Consonant | Sequence | root |
| ckv1 | Consonant | Sequence | root |
| sst1 | Tone | Positional | root |
| kzj1 | Consonant | Sequence | root |
| nnb1 | Vowel | Inventory | morpheme |
| nym1 | Consonant | Sequence | root |
| nym2 | Vowel | Inventory | N affix |
| xom1 | Vowel | Inventory | morpheme |
| kpy1 | Consonant | Positional | morpheme |
| kye1 | Vowel | Inventory | morpheme |
| bfa1 | Consonant | Positional | root |
| lkt1 | Syllable | Inventory | root proper |
| las1 | Vowel | Inventory | morpheme |
| lag1 | Vowel | Inventory | morpheme |
| laj1 | Consonant | Positional | root |


| Index | Substance | Type | Domain |
| :---: | :---: | :---: | :---: |
| laj2 | Consonant | Positional | root proper |
| laj3 | Consonant | Positional | root proper |
| laj4 | Vowel | Inventory | affix |
| laj5 | Vowel | Inventory | root proper |
| les1 | Vowel | Inventory | morpheme |
| bod1 | Vowel | Inventory | morpheme |
| kmk1 | Consonant | Sequence | root |
| llq1 | Consonant | Sequence | root |
| dni1 | Consonant | Positional | morpheme |
| dni2 | Consonant | Positional | morpheme |
| dni3 | Consonant | Sequence | morpheme |
| dni4 | Consonant | Sequence | morpheme |
| dni5 | Consonant | Sequence | morpheme |
| dni6 | Vowel | Sequence | morpheme |
| dni7 | Vowel | Sequence | morpheme |
| dni8 | Vowel | Sequence | morpheme |
| dni9 | Vowel | Sequence | morpheme |
| mal1 | Vowel | Positional | root |
| mal2 | Consonant | Sequence | root |
| mgq1 | Vowel | Inventory | morpheme |
| mdk1 | Vowel | Inventory | morpheme |
| arn1 | Consonant | Minimality/Maximality | root |
| vma1 | Consonant | Sequence | morpheme |
| mdm1 | Vowel | Inventory | morpheme |
| tay1 | Consonant | Sequence | root |
| jkr1 | Vowel | Sequence | morpheme |
| jkr2 | Syllable | Inventory | morpheme |
| mnb1 | Consonant | Sequence | root |
| mnb2 | Consonant | Sequence | root |
| mnb3 | Consonant | Sequence | root |
| mnb4 | Consonant | Sequence | root |
| mnb5 | Consonant | Sequence | root |
| mnb6 | Consonant | Sequence | root |
| mqu1 | Vowel | Inventory | morpheme |
| sur1 | Consonant | Positional | morpheme |
| sur2 | Vowel | Positional | morpheme |
| sur3 | Consonant | Positional | morpheme |
| sur 4 | Vowel | Sequence | morpheme |


| Index | Substance | Type | Domain |
| :---: | :---: | :---: | :---: |
| acx1 | Consonant | Sequence | root |
| acx2 | Consonant | Inventory | affix |
| mxh1 | Vowel | Inventory | morpheme |
| jup1 | Consonant | Positional | morpheme |
| naw1 | Vowel | Inventory | morpheme |
| niy1 | Consonant+Vowel | Sequence | morpheme |
| niy2 | Consonant+Vowel | Sequence | morpheme |
| niy3 | Vowel | Positional | root |
| niy4 | Consonant | Positional | V root |
| niy5 | Consonant | Positional | V root |
| niy6 | Tone | Positional | V root |
| nko1 | Consonant | Positional | morpheme |
| ace1 | Consonant | Sequence | root |
| nso1 | Vowel | Inventory | morpheme |
| nna1 | Vowel | Positional | N root |
| nna2 | Vowel | Positional | V root |
| nna3 | Consonant | Sequence | morpheme |
| nna4 | Consonant | Sequence | morpheme |
| nzb1 | Vowel | Inventory | root |
| opm1 | Consonant | Sequence | morpheme |
| $\operatorname{lgn} 1$ | Vowel | Inventory | morpheme |
| pwn1 | Consonant | Sequence | root |
| pau1 | Consonant | Sequence | root |
| lkr1 | Vowel | Inventory | affix |
| spa1 | Vowel | Inventory | morpheme |
| cmn1 | Consonant | Positional | root |
| cmn2 | Consonant | Inventory | root |
| fuc1 | Vowel | Inventory | morpheme |
| puu1 | Vowel | Inventory | morpheme |
| pyu1 | Consonant | Sequence | root |
| rus1 | Vowel | Sequence | morpheme |
| bps1 | Consonant | Sequence | root |
| mbs1 | Consonant | Sequence | root |
| trv1 | Consonant | Sequence | root |
| sza1 | Consonant | Positional | morpheme |
| sza2 | Vowel | Positional | root |
| sot1 | Vowel | Inventory | morpheme |
| tsn1 | Vowel | Inventory | morpheme |


| Index | Substance | Type | Domain |
| :---: | :---: | :---: | :---: |
| kca1 | Vowel | Positional | morpheme |
| erg1 | Syllable | Minimality/Maximality | root |
| erg2 | Consonant+Vowel | Sequence | root |
| erg3 | Consonant | Positional | V root |
| erg4 | Consonant | Positional | V root |
| erg5 | Consonant | Positional | root |
| erg6 | Consonant | Positional | N root |
| erg7 | Consonant | Sequence | root |
| erg8 | Consonant | Sequence | root |
| tgl1 | Consonant | Sequence | root |
| tlj1 | Vowel | Inventory | morpheme |
| tpj1 | Consonant | Sequence | root |
| ssf1 | Consonant | Sequence | root |
| tih1 | Consonant | Sequence | root |
| suq1 | Consonant | Positional | affix |
| bbc1 | Consonant | Sequence | root |
| dto1 | Consonant | Sequence | root |
| dto2 | Consonant | Positional | root |
| dto3 | Consonant | Positional | root |
| dto4 | Vowel | Positional | root |
| dto5 | Tone | Minimality/Maximality | root |
| dto6 | Tone | Minimality/Maximality | root |
| dto7 | Tone | Positional | N root |
| tri1 | Consonant+Vowel | Sequence | morpheme |
| tso1 | Vowel | Inventory | morpheme |
| tur1 | Consonant+Vowel | Sequence | root |
| tyv1 | Consonant | Sequence | root |
| ulw1 | Consonant | Inventory | morpheme |
| uur1 | Syllable | Minimality/Maximality | root proper |
| uur2 | Consonant | Positional | N root |
| uur3 | Consonant | Positional | V root |
| uur4 | Consonant | Positional | root |
| ghl1 | Vowel | Inventory | morpheme |
| wmb1 | Consonant | Positional | N root proper |
| wmb2 | Consonant | Positional | V root |
| $\operatorname{lgg} 1$ | Vowel | Inventory | morpheme |
| yev1 | Consonant | Positional | root |
| yev2 | Syllable | Inventory | root |


| Index | Substance | Type | Domain |
| :--- | :--- | :--- | :--- |
| ctz1 | Vowel | Positional | root |
| zne1 | Vowel | Inventory | morpheme |
| zul1 | Vowel | Inventory | morpheme |

Table 5: Infixation Sample

| Index | Infixation | InfixInfo | InfixInteraction |
| :--- | :--- | :--- | :--- |
| tbw1 | Y | -um- actor focus | Muna-type |
| akk1 | Y | -ta- reflexive | None |
| akk2 | Y | -ta- reflexive | Muna-type |
| arb1 | Y | -t- reflexive | Hebrew-type |
| heb1 | Y | -t- reflexive | Hebrew-type |
| huz1 | Y | -á- pluractionality | Hebrew-type |
| xnb1 | Y | -um- actor voice | Muna-type |
| pam1 | Y | -um-actor focus | Muna-type |
| ckv1 | Y | -um- agent focus | Muna-type |
| kzj1 | Y | -um- nominative focus | Muna-type |
| lkt1 | Y | -ung(k)- 1.plural subject | None |
| kmk1 | Y | -um- actor focus, inchoative | Muna-type |
| llq1 | Y | -um- future | Muna-type |
| dni1 | Y | -si- perfect | none |
| dni2 | Y | -si- perfect | none |
| dni3 | Y | -si- perfect | none |
| dni4 | Y | -si- perfect | none |
| dni5 | Y | -si- perfect | none |
| dni6 | Y | -si- perfect | none |
| dni7 | Y | -si- perfect | none |
| dni8 | Y | -si- perfect | none |
| dni9 | Y | -si- perfect | none |


| Index | Infixation | InfixInfo | InfixInteraction |
| :--- | :--- | :--- | :--- |
| tay1 | Y | -um- actor focus | Muna-type |
| mnb1 | Y | -um- irrealis | None |
| mnb2 | Y | -um- irrealis | None |
| mnb3 | Y | -um- irrealis | None |
| mnb4 | Y | -um- irrealis | Muna-type |
| mnb5 | Y | -um- irrealis | None |
| mnb6 | Y | -um- irrealis | None |
| sur1 | Y | -a- verbal plural | None |
| sur2 | Y | -a- verbal plural | None |
| sur3 | Y | -a- verbal plural | None |
| sur4 | Y | -a- verbal plural | None |
| acx1 | Y | -t- reflexive | Hebrew-type |
| acx2 | Y | -t- reflexive | Hebrew-type |
| ace1 | Y | -um- intransitive | Muna-type |
| pwn1 | Y | -m- actor focus | Muna-type |
| pau1 | Y | $-m-$ verb marker | Muna-type |
| cmn1 | Y | -l- diminutive | Hebrew-type |
| cmn2 | Y | -l- diminutive | Hebrew-type |
| pyu1 | Y | -em- actor voice | Muna-type |
| bps1 | Y | -m- actor focus | Muna-type |
| mbs1 | Y | -om- future | Muna-type |
| trv1 | Y | -um- actor focus | Muna-type |
| sza1 | Y | -r- causative, infixing reduplication, intransitive | none, Hebrew-type |
| sza2 | Y | -r- causative, infixing reduplication, intransitive | none, Hebrew-type |


| Index | Infixation | InfixInfo | InfixInteraction |
| :--- | :--- | :--- | :--- |
| tgl1 | Y | -um- actor focus | Hebrew-type |
| ssf1 | Y | -um- actor focus | Muna-type |
| tih1 | Y | -um- future | Muna-type |
| bbc1 | Y | -um- completive | Muna-type |
| ulw1 | Y | -ka- construct state | None |
| yev1 | Y | -n- singular masculine object, reduced, -m- imperfective | Hebrew-type |
| yev2 | Y | -ie- irrealis | Hebrew-type |

Table 6: Asymmetric non-trigger sample

| Index | Vowel Harmony | Haromonizing Feature | Asymmetric non-trigger |
| :---: | :---: | :---: | :---: |
| aha1 | Y | [+ATR] | $\stackrel{\text { a }}{ }$ |
| yor5 | Y | [-ATR] | I, U |
| rwm1 | Y | [+ATR] | e,o |
| gwx1 | Y | [+ATR] | a |
| asm1 | Y | [+ATR] | e,o |
| avi1 | Y | [-ATR] | I, ${ }^{\text {e }}$ |
| avu1 | Y | [+ATR] | e,o |
| bdh1 | Y | [+ATR] | e,o |
| bct1 | Y | [+ATR] | e,o |
| bdi1 | Y | [+ATR] | e,o |
| sro1 | Y | [+ATR] | e,o |
| ncu1 | Y | [+ATR] | $\underset{\sim}{\text { a }}$ |
| did1 | Y | [+ATR] | $\underset{\sim}{\text { a }}$ |
| mhr1 | Y | [back] | æ |
| etu11 | Y | [+ATR] | e,o |
| yor2 | Y | [-ATR] | I, U |
| fvr1 | Y | [+ATR] | e,o, ${ }_{\text {a }}$ |
| acd1 | Y | [+ATR] | $\underset{\sim}{\text { a }}$ |
| toh1 | Y | [+ATR] | e,o |
| g.jn1 | Y | [+ATR] | $\underset{+}{\text { a }}$ |
| rub1 | Y | [+ATR] | e,o |
| yor4 | Y | [-ATR] | I, U |
| yor1 | Y | [-ATR] | I, ${ }^{\text {d }}$ |
| yor3 | Y | [-ATR] | I, U |
| jum1 | Y | [+ATR] | e,o |
| ksp1 | Y | [-ATR] | $\varepsilon$ |
| keol | Y | [+ATR] | e,o |
| nnb1 | Y | [+ATR] | e,o,a |
| xom1 | Y | [+ATR] | e,o,a |
| kye1 | Y | [+ATR] | $\underset{\sim}{\text { a }}$ |
| las1 | Y | [+ATR] | $\underset{+}{\text { a }}$ |
| lag1 | Y | [+ATR] | e,o |
| les1 | Y | [+ATR] | e,o |
| bod1 | Y | [+high] | ¢ |
| mgq1 | Y | [+ATR] | e,o |
| mdk1 | Y | [+ATR] | e,o |


| Index | Vowel Harmony | Haromonizing Feature | Asymmetric non-trigger |
| :---: | :---: | :---: | :---: |
| mdm1 | Y | [+ATR] | e,o |
| mqu1 | Y | [+ATR] | e,o |
| mxh1 | Y | [+ATR] | e,o |
| naw1 | Y | [+ATR] | $\underset{+}{\text { a }}$ |
| nso1 | Y | [+raised] | ${ }_{1}^{1},{ }_{\sim}^{\text {U }}$ |
| nzb1 | Y | [+ATR] | e,o |
| $\operatorname{lgn} 1$ | Y | [+ATR] | e,o,a |
| lkr1 | Y | [ $\pm$ ATR] | e,o,, , u |
| spa1 | Y | [-ATR] | $\bigcirc$, I |
| fuc1 | Y | [+ATR] | e,o |
| puu1 | Y | [+ATR] | e,o |
| sot1 | Y | [+raised] |  |
| tsn1 | Y | [+raised] | ${ }_{1}$, U |
| kca1 | Y | [back] | æ,y,u |
| $t_{\text {tlj }} 1$ | Y | [+ATR] | e,o,a |
| tso1 | Y | [+ATR] | e,o |
| ghl1 | Y | [+ATR] | e,o |
| $\operatorname{lgg} 1$ | Y | [+ATR] | $\underset{+}{\text { a }}$ |
| zne1 | Y | [-ATR] | e,o |
| zul1 | Y | [-ATR] | e,o |

Table 7: List of sources in DoCoMD

| Index | Source |
| :--- | :--- |
| tbw1 | Zuraw \& Lu (2009:200) |
| kjq1 | Miller (1965:24,42) |
| aha1 | Casali (2003:363) |
| aka1 | Clements (1985:64) |
| akk1 | Lubowicz (2010), Yip (1988:72) |
| akk2 | Lubowicz (2010:15), Yip (1988) |
| yor5 | Casali (2003:338) |
| rwm1 | Rolle et al. (2020), Kutsch-Lojenga (2008:66) |
| gwx1 | Casali (2003:363) |
| arb1 | McCarthy (1981:396) |
| asm1 | Mahanta (2008:113) |
| avi1 | Rolle et al. (2020) |
| avu1 | Rolle et al. (2020) |
| bdh1 | Rolle et al. (2020) |
| bct1 | Rolle et al. (2020) |
| bdi1 | Rolle et al. (2020), Casali (2003), Andersen (1999:4) |
| sro1 | Torres-Tamarit et al. (2017) |
| yue1 | Yip (1988:82) |
| cao1 | Tallman (2018:72) |
| cao2 | Tallman (2018:73) |
| ncu1 | Casali (2003:363) |
| njo1 | Temsunungsang (2009:103) |
| yao1 | $\boldsymbol{v}$ |
| yao2 | Ngunga (1997:22) |
| yao3 | Ngunga (1997:17) |
| com1 | Ngunga (1997:58) |
| tfn1 | Charney (1993:23) |
| tfn2 | Lovick (2020:45) |
| tfn3 | Lovick (2020:63) |
| tfn4 | Lovick (2020:75) |
| tfn4 | Lovick (2020:119) |
| des1 | Lovick (2020:120) |
| des2 | Silva (2012:54) |
| did1 | Silva (2012:42) |
| dji1 | Casali (2003:321), van der Hulst (2018:295) |


| Index | PBase | Source |
| :---: | :---: | :---: |
| nld1 |  | Booij (2011:2052) |
| mhr1 |  | Vaysman (2009:61) |
| etu1 | $\checkmark$ | Watters (1981:42) |
| etu2 | $\checkmark$ | Watters (1981:43) |
| etu3 |  | Watters (1981:43) |
| etu4 |  | Watters (1981:44) |
| etu5 |  | Watters (1981:44) |
| etu6 |  | Watters (1981:44) |
| etu7 | $\checkmark$ | Watters (1981:48) |
| etu8 |  | Watters (1981:48) |
| etu9 | $\checkmark$ | Watters (1981:48) |
| etu10 |  | Watters (1981:54) |
| etu11 |  | Watters (1981:54) |
| etu11 |  | Casali (2003:338) |
| yor2 |  | Rolle et al. (2020), Casali (2003:325) |
| eng1 |  | Booij (2011:2052) |
| fmp1 | $\checkmark$ | Hyman (1972:31) |
| fmp2 | $\checkmark$ | Hyman (1972:30) |
| fmp3 |  | Hyman (1972:40) |
| fmp4 |  | Hyman (1972:40) |
| fmp5 |  | Hyman (1972:41) |
| fmp6 |  | Hyman (1972:41) |
| fmp 7 |  | Hyman (1972:41) |
| fvr1 |  | Casali (2003:336), van der Hulst (2018:298) |
| gaa1 | $\checkmark$ | Zimmerman (1858:6) |
| gaa2 | $\checkmark$ | Zimmerman (1858:6) |
| gaa3 | $\checkmark$ | Zimmerman (1858:6) |
| cab1 |  | Haurholm-Larsen (2016:24) |
| deu1 |  | Golston \& Wiese (1998:174) |
| deu2 |  | Golston \& Wiese (1998:177) |
| acd1 |  | Casali (2003:363) |
| toh1 |  | Casali (2003:338) |
| gis1 |  | Shay (2021:25) |
| gjn1 |  | Casali (2003:363) |
| gni1 | $\checkmark$ | McGregor (1990:70) |
| gni2 | $\checkmark$ | McGregor (1990:77) |
| gni3 | $\checkmark$ | McGregor (1990:77) |
| rub1 |  | Rolle et al. (2020), Kutsch-Lojenga (2008:66) |


| Index | PBase | Source |
| :---: | :---: | :---: |
| heb1 |  | McCarthy (1986:209) |
| qub1 |  | Weber (1989:455) |
| hun1 |  | Siptár \& Törkenczy (2000:143) |
| hun2 |  | Siptár \& Törkenczy (2000:145) |
| hun3 |  | Siptár \& Törkenczy (2000:145) |
| hun4 |  | Siptár \& Törkenczy (2000:152) |
| huz1 |  | van den Berg (1995:27), Kalin (2021:3) |
| yor4 |  | Rolle et al. (2020), Casali (2003:325) |
| yor1 |  | Rolle et al. (2020), Casali (2003:325) |
| ind1 | $\checkmark$ | Lapowila (1981:72) |
| yor3 |  | Rolle et al. (2020), Casali (2003:325) |
| ita1 |  | Itô \& Mester (2021:4) |
| mks1 |  | Carroll (2015:37) |
| mks2 |  | Carroll (2015:57) |
| mks3 |  | Carroll (2015:57) |
| nca1 |  | Minter (2009:11) |
| jav1 |  | Mester (1987:98) |
| jav2 |  | Mester (1987:98) |
| jum1 |  | Rolle et al. (2020), Andersen (2004:136) |
| ksp1 |  | Casali (2003:325), van der Hulst (2018:300) |
| keo1 |  | Rolle et al. (2020), Onziga \& Gilley (2012:7) |
| kis1 |  | Arsenault \& Kochetov (2019) |
| xnb1 |  | Zuraw \& Lu (2009:200) |
| pam1 |  | Zuraw \& Lu (2009:200) |
| ckv1 |  | Li \& Tsuchida (2006:4), Zuraw \& Lu (2009:200) |
| sst1 |  | Rarrick (2017:99) |
| kzj1 |  | Zuraw \& Lu (2009:200) |
| nnb1 |  | Casali (2003:351), van der Hulst (2018:265) |
| nym1 | $\checkmark$ | Maganga \& Schadeberg (1992:23) |
| nym2 | $\checkmark$ | Maganga \& Schadeberg (1992:26) |
| xom1 |  | Rolle et al. (2020), Otero (2015:213) |
| kpy1 |  | Abramovitz (2019:13) |
| kye1 |  | Casali (2003:363) |
| bfa1 | $\checkmark$ | Cohen (2000:52) |
| lkt1 |  | Albright (2004:2), Albright (2000) |
| las1 |  | Casali (2003), Ourso (1989:175) |
| lag1 |  | Rolle et al. (2020), Casali (2003:351), van der Hulst (2018:298) |
| laj1 |  | Noonan (1992:7) |


| Index | PBase | Source |
| :---: | :---: | :---: |
| laj2 |  | Noonan (1992:15) |
| laj3 | $\checkmark$ | Noonan (1992:8) |
| laj4 | $\checkmark$ | Noonan (1992:30) |
| laj5 |  | Noonan (1992:30) |
| les1 |  | Rolle et al. (2020), Casali (2003:336) |
| bod1 |  | van der Hulst (2018:272) |
| kmk1 |  | Zuraw \& Lu (2009:210) |
| llq1 |  | Zuraw \& Lu (2009:200) |
| dni1 | $\checkmark$ | Bromley (1961:62), Bromley (1981:24) |
| dni2 | $\checkmark$ | Bromley (1961:62), Bromley (1981:24) |
| dni3 |  | Bromley (1961:64), Bromley (1981:24) |
| dni4 |  | Bromley (1961:64), Bromley (1981:24) |
| dni5 |  | Bromley (1961:64), Bromley (1981:24) |
| dni6 |  | Bromley (1961:64), Bromley (1981:24) |
| dni7 |  | Bromley (1961:67), Bromley (1981:24) |
| dni8 |  | Bromley (1961:64), Bromley (1981:24) |
| dni9 |  | Bromley (1961:66), Bromley (1981:24) |
| mal1 | $\checkmark$ | Asher \& Kumari (1997:428) |
| mal2 |  | Mohanan (1989:600) |
| mgq1 |  | Rolle et al. (2020), Kutsch-Lojenga (2008:66) |
| mdk1 |  | Rolle et al. (2020) |
| arn1 |  | Smeets (2008:37) |
| vma1 | $\checkmark$ | Dench (1995:32) |
| mdm1 |  | Casali (2003:322), van der Hulst (2018:295) |
| tay1 |  | Zuraw \& Lu (2009:203) |
| jkr1 |  | Modi (2017:95) |
| jkr2 |  | Modi (2017:95) |
| mnb1 | $\checkmark$ | van den Berg (1989:30) |
| mnb2 | $\checkmark$ | van den Berg (1989:30) |
| mnb3 | $\checkmark$ | van den Berg (1989:30) |
| mnb4 | $\checkmark$ | van den Berg (1989:31) |
| mnb5 | $\checkmark$ | van den Berg (1989:31) |
| mnb6 | $\checkmark$ | van den Berg (1989:31) |
| mqu1 |  | Rolle et al. (2020), Stirtz (2014:6) |
| sur1 | $\checkmark$ | Frajzyngier (1993:3) |
| sur2 |  | Frajzyngier (1993:3) |
| sur3 | $\checkmark$ | Frajzyngier (1993:3) |
| sur4 |  | Frajzyngier (1993:4) |


| Index | PBase | Source |
| :---: | :---: | :---: |
| acx1 |  | Glover (1988:42) |
| acx2 | $\checkmark$ | Glover (1988:42) |
| mxh1 |  | Rolle et al. (2020), Casali (2003:336) |
| jup1 |  | Epps (2008:46) |
| naw1 |  | Casali (2003:363) |
| niy1 | $\checkmark$ | Kutsch Lojenga (1994:35) |
| niy2 | $\checkmark$ | Kutsch Lojenga (1994:36) |
| niy3 |  | Kutsch Lojenga (1994:60) |
| niy4 |  | Kutsch Lojenga (1994:61) |
| niy5 |  | Kutsch Lojenga (1994:61) |
| niy6 |  | Kutsch Lojenga (1994:61) |
| nko1 |  | Peacock (2007:20) |
| ace1 |  | Zuraw \& Lu (2009:200) |
| nso1 |  | van der Hulst (2018:269) |
| nna1 |  | O'Grady (1963:29) |
| nna2 | $\checkmark$ | O'Grady (1963:29) |
| nna3 | $\checkmark$ | O'Grady (1963:29) |
| nna4 |  | O'Grady (1963:29) |
| nzb1 |  | van der Hulst (2018:264) |
| opm1 |  | Loughnane (2009:68) |
| $\operatorname{lgn} 1$ |  | Rolle et al. (2020), Mellese (2017:35) |
| pwn1 |  | Zuraw \& Lu (2009:200) |
| pau1 |  | Zuraw \& Lu (2009:200), Lubowicz (2010:4) |
| lkr1 |  | Andersen (1989:3) |
| spa1 |  | van der Hulst (2018:435) |
| cmn1 |  | Yu (2004:40) |
| cmn1 |  | Yu (2004:40) |
| fuc1 |  | Casali (2003:321), van der Hulst (2018:267) |
| puu1 |  | van der Hulst (2018:267) |
| pyu1 |  | Teng (2008:25), Zuraw \& Lu (2009:200) |
| rus1 |  | Halle (1959:30) |
| bps1 |  | Zuraw \& Lu (2009:200) |
| mbs1 |  | Zuraw \& Lu (2009:200) |
| trv1 |  | Zuraw \& Lu (2009:200) |
| szal |  | Kruspe (2004:52) |
| sza2 |  | Kruspe (2004:53) |
| sot1 |  | van der Hulst (2018:269) |
| tsn1 |  | van der Hulst (2018:269) |


| Index | PBase | Source |
| :---: | :---: | :---: |
| kca1 |  | van der Hulst (2018:178), Vaysman (2009:6) |
| erg1 |  | Crowley (1998:18) |
| erg2 |  | Crowley (1998:18) |
| erg3 | $\checkmark$ | Crowley (1998:21) |
| erg4 |  | Crowley (1998:21) |
| erg5 |  | Crowley (1998:21) |
| erg6 | $\checkmark$ | Crowley (1998:20) |
| erg7 | $\checkmark$ | Crowley (1998:23) |
| erg8 | $\checkmark$ | Crowley (1998:23) |
| tgl1 |  | Zuraw \& Lu (2009:200) |
| tlj1 |  | Rolle et al. (2020), Casali (2003) |
| tpj1 |  | González (2005:62) |
| ssf1 |  | Blust (2003:23), Zuraw \& Lu (2009:200) |
| tih1 |  | Zuraw \& Lu (2009:200) |
| suq1 | $\checkmark$ | Bryant (1999:30) |
| bbc1 |  | Zuraw \& Lu (2009:200) |
| dto1 |  | McPherson (2013:17) |
| dto2 |  | McPherson (2013:40) |
| dto3 |  | McPherson (2013:40) |
| dto4 |  | McPherson (2013:42) |
| dto5 |  | McPherson (2013:75) |
| dto6 |  | McPherson (2013:75) |
| dto 7 |  | McPherson (2013:77) |
| tri1 |  | Meira (1999:58) |
| tsol |  | Casali (2003:321) |
| tur1 |  | Lees (1966:35) |
| tyv |  | Voinov (2016:10) |
| ulw1 |  | Green (1999:38) |
| uur1 |  | Crowley (1999:116) |
| uur2 |  | Crowley (1999:118) |
| uur3 |  | Crowley (1999:119) |
| uur4 |  | Crowley (1999:119) |
| ghl1 |  | Rolle et al. (2020) |
| wmb1 |  | Nordlinger (1998:25) |
| wmb2 |  | Nordlinger (1998:25) |
| $\operatorname{lgg} 1$ |  | Casali (2003), van der Hulst (2018:295), Andersen (1986:57) |
| yev1 |  | Wilson (2017:70,352) |
| yev2 |  | Wilson (2017:70) |


| Index | PBase | Source |
| :--- | :--- | :--- |
| ctz1 | Villard $(2015: 139)$ |  |
| zne1 | Casali $(2003: 338)$, van der Hulst $(2018: 298)$ |  |
| zul1 | Casali $(2003: 338)$, van der Hulst $(2018: 266)$ |  |


[^0]:    ${ }^{1}$ An additional property originally proposed for MSCs is that they are feature-filling (Halle 1959; Stanley 1967; Kisseberth \& Kenstowicz 1977) or structure preserving. This is largely an assumption dependent on a certain theory of input underspecification and will not be discussed any further, cf. Inkelas (1995) for a theory of input underspecification in OT.

[^1]:    ${ }^{2}$ Here and in the following, I mark ungrammatical forms that do not result from a misapplication of a phonological process with **.

[^2]:    ${ }^{3}$ Similar arguments have been made by van Oostendorp (2001) for $/ \mathrm{gg} /$ in Dutch and German dialects (Germanic; Germany, Netherlands), by Vaysman (2002) for consonant gradation in Nganasan (Uralic, Russia), by Jones (2014:196) for tone in Kinande (Atlantic-Congo; Democratic Republic of Congo, Uganda), and by Staroverov (2020) for $/ \mathrm{k} /$ in Tundra Nenets (Uralic, Russia). I will not discuss these patterns or Finnish assibilation for that matter in the remainder of this thesis and instead focus on patterns that pose more concrete challenges to Richness of the Base.

[^3]:    ${ }^{4}$ A similar idea has been proposed by Kiparsky (1982), where instead of CoURs, generalizations are assumed to be true at the stem level.

[^4]:    ${ }^{5}$ There is an active debate in the literature if there is any conclusive evidence in favor of the need for recursive prosodic structure. I do not claim to add to this debate. Instead, I simply point out that recursive domains allow us to describe monomorphemic affix domains. For recent discussions on this topic

[^5]:    see Downing \& Kadenge (2020), Itô \& Mester (2021), and Miller \& Sande (2021).

[^6]:    ${ }^{6}$ Morpheme here refers to a minimal linguistic form or a morph in the sense of Haspelmath (2020).
    ${ }^{7}$ Such a domain is plausibly present in Nimboran (Nimboranic, Indonesia), as shown in Inkelas (1993).
    ${ }^{8}$ One could of course include another constraint domain that encompasses only material dominated by a minimal prosodic word. Such a domain would not add anything new, since it would pick out morphemes in coordinative structures and roots in adjunctive structures. Morphemes are already the domain of $\omega, \mathrm{DD}$

[^7]:    and roots are already in the domain of $\sqrt{\omega}$.
    ${ }^{9}$ There is evidence that only some category-changing affixes are treated as roots by phonology (cf. Creemers et al. 2018). Additionally, roots with a category-feature might appear as complex objects in morphosyntactic structures under a Distributed Morphology approach to the morphology-phonology interface (Marantz 1997; Embick 2003; Harley 2005; Embick \& Marantz 2008). This does not affect the

[^8]:    ${ }^{10}$ It is not crucial here if the affix is itself dominated by its own prosodic word.

[^9]:    ${ }^{1}$ This is a slight divergence from Bickel (2008), who would include the data point under the name of the language family. Additionally, I only excluded languages whenever all values for a given family were identical, instead of determining for each group the statistical likeliness of genealogical skewing. This was mostly done due to the already small size of the database.

[^10]:    ${ }^{2}$ The other category covers all top-level families of which only one language is included in the database. This includes one languages from the following top-level families: Araucanian, Arawakan, Athabaskan-Eyak-Tlingit, Austroasiatic, Bunaban, Cariban, Chukotko-Kamchatkan, Dogon, Dravidian, Furan, Keresan, Mirndi, Misumalpan, Naduhup, Nakh-Dagestanian, Nubian, Nuclear Torricelli, Pano-Tacanan, Quechuan, Sahaptian, Siouan, Tucanoan, Tupian, Turkic, and Uto-Aztecan.

[^11]:    ${ }^{3}$ See van Oostendorp (2014) for a theoretical implementation of such patterns inside Optimality Theory.

[^12]:    ${ }^{4}$ Here and in the following, notations have been adjusted to IPA. Tones are always marked with diacritics above the vowel. Unmarked low or mid tones in the source are explicitly marked.

[^13]:    ${ }^{5}$ It seems that sequential constraints do not make reference to derived properties as robustly as the other constraint types. I have no explanation for this and leave the question open for future research.

[^14]:    ${ }^{6}$ Apocope never applies in lexical roots (Albright 2004).

[^15]:    ${ }^{7}$ Unfortunately, no clear examples of prefixation are found in McGregor (1990) and therefor no example of a non-word-initial root or a prefix starting with [r] can be provided here.

[^16]:    ${ }^{8}$ High back vowels are banned from this position for independent reasons. Therefore, only front high vowels are shown here.

[^17]:    ${ }^{9}$ In the following, I only report the most significant interactions with $\mathrm{p}<0.01$.

[^18]:    ${ }^{10}$ Note that these constraints interact with vowel harmony (cf. chapter 4) but are never the result of harmony. Therefore their presence is not trivially dependent on the presence of vowel harmony.

[^19]:    ${ }^{11}$ Direct domination only refers to material that is dominated by a prosodic word without any intervening prosodic words. See chapter 1 for more details. Note also that the structure provided is only one possible structure. Languages might differ in what prosodic domains are relevant. Additionally, affixes might differ in their status, as exemplified by the case study of Chungli Ao in chapter 5.

[^20]:    ${ }^{12}$ Root material is marked with a subscript R . The attested non-optimal candidate is marked with $\sharp$, the unattested predicted output with $\leftarrow$.

[^21]:    ${ }^{1}$ The actual constraint bans homorganic nasals and plosives (prenasalized or not) in the first and second consonant position of a root. Other relevant constraints exclude roots with contra-voiced homoorganic plosives (and contravoiced homoorganic obstruents if velar or labial). Similarly, prenasalized plosives cannot cooccur in both position simultaneously.
    ${ }^{2}$ Carter (2000) analyzes this as morpheme-specific left-alignment of a featural affix [labial][voiced][nasal] instead, where these features covertly dock to any labial preceding the infix in this context. This misses the generalization that the constraint also holds over root domains.

[^22]:    ${ }^{3}$ There are other restrictions that trigger the same repair mechanism, namely the ban on coocurring prenasalized consonants or nasals.

[^23]:    ${ }^{4}$ I assume that these are two exponents of the same morpheme, i.e. with different positional requirements in phonology but the same morphological affiliation.
    ${ }^{5}$ All Hebrew data in this section are from Ezer Rasin (p.c.), who is a native speaker of Modern Hebrew and a trained phonologist.

[^24]:    ${ }^{6}$ This pattern has been described as metathesis, but the distinction does not bear on the argument here. The only crucial point is that the position of $\langle\mathrm{t}\rangle$ is outside of the root in the input and inside the root on the surface. A potential argument against a phonologically general metathesis pattern is the absence of metathesis in other morphological contexts as in khad-ts'dadí 'one-sided' and tát-sidrá 'sub-order'.

[^25]:    ${ }^{7}$ An alternative characterization of the data - based on syllable structure and suggested to me by Juliet Stanton (p.c.) assumes that the roots cannot include two identical onsets in adjacent syllables (cf. also Cohn 1992; Rose \& Walker 2004). This would predict that the forms in (14) are licit, contrary to fact.

[^26]:    ${ }^{8}$ If one were to assume that the same BM-repairs apply to the root+infix and to the root domain, the surface forms would end up looking similar to infixed forms, even if they did not include any infix. I will try to avoid this confusion in order to not divert from the main point here.

[^27]:    ${ }^{9}$ This approach is therefore incompatible with the serial approach to infixation in Kalin (2021), even though both approaches are based on very similar data.

[^28]:    ${ }^{10}$ This does not make any statement about the boundness of infixes. See also chapter 1 for the assumption that all kinds of affixes can potentially be dominated by their own prosodic word.

[^29]:    ${ }^{11}$ In the following tableaux, infix material is marked in blue. I also abstract away from syllable and foot structure. See chapter 5 for a discussion.
    ${ }^{12} \mathrm{I}$ do not discuss candidates here with different repair mechanisms. A candidate [momi] would be a possible output under the current ranking, with the initial $/ \mathrm{m} /$ either being part of the infix or of the root. Their exclusion depends on the exact nature of the nasal accretion process, which is not directly relevant here.

[^30]:    ${ }^{13}$ For the sake of exposition, the candidate in (25-d) dissimilates the infix instead of deleting it. This is different from candidate (23-d).

[^31]:    ${ }^{14}$ The fixed ranking can thus explain an empirical tendency in the languages of the world to adhere to the Strong Domain Hypothesis. It does allow for exceptions by structure though, which is in accordance with the exceptions that have been provided in the literature, see section 3.5 for more discussion.

[^32]:    ${ }^{15}$ Similarly, Peperkamp (1996) argues that Align constraints always exclude certain prosodic structures for clitics and McCarthy \& Prince (1993) derive a related generalization in a Generalized Alignment framework for Dakota infixation.
    ${ }^{16}$ These structures are similar to the proposal made by Lubowicz (2010) where infixes can be absorbed

[^33]:    into other morphemes. Note that Lubowicz (2010) explicitly refers to morphological structure being manipulated by phonological operations. This is incompatible with a modular approach to grammar, cf. also van Oostendorp (2007) for a discussion.

[^34]:    ${ }^{17}$ In fact, Hebrew ${ }^{\prime}$ could be derived if infixation and repair rules were disjunctively ordered (cf. e.g. Chomsky \& Halle 1968:30, Anderson 1986). A MS-Rule and a P-rule could then be blocked if infixation applied. Since disjunctive ordering was never intended to be used for the interaction of infixation and phonology, I will assume that the most charitable representation of such a theory does not predict Hebrew'.

[^35]:    ${ }^{18}$ Restricting the Strong Domain Hypothesis to certain processes, faithfulness constraints, or markedness constraints, e.g. dissimilation/OCP (Suzuki 1998) is not a viable option either. As shown in the course of this chapter, infixation interacts with dissimilation/sequential constraints (Muna, Hebrew), maximality constraints/deletion (Hunzib), and syllable structure (Semelai, Yeri).

[^36]:    ${ }^{19}$ The patterns of Hebrew and Muna can be derived by ordering infixation at a later stratum after a stratum with a $\mathrm{M} \gg \mathrm{F}$ ranking where the repair has already taken place. For Hebrew, infixation has to happen at a stratum where the faithfulness constraint is ranked above the markedness constraint, in order for the infix to not join the domain of the $\mathrm{OCP}(\mathrm{C})$ constraint. For Muna, the ranking has to be reversed. The second stratum keeps the $\mathrm{M} \gg \mathrm{F}$ and the infix joins the domain of the BM constraint.

[^37]:    ${ }^{20}$ See, however, Kiparsky (2015:33), who argues that the Strong Domain Hypothesis is not easily transferable to Stratal OT and cannot be maintained in its simplest implementation for empirical reasons. Kiparsky (2015:16) exemplifies this with data from Mesopotamian and Syrian-Palestinian Arabic where final geminates are degeminated at the word-level but not at the stem level as evident from stress assignment.
    ${ }^{21}$ Such an implementation makes it difficult to single out certain markedness constraints or markedness constraint types that need to be switched off. This problem could be solved by instead restricting reranking to demotion of markedness constraints, contra Itô \& Mester (1999).

[^38]:    ${ }^{22}$ The pattern derived here is instead a Muna pattern.

[^39]:    ${ }^{23}$ If infixes were representationally special in that they were enclosed in root boundary marker ' + ', the resulting string representations would result in a Hebrew' pattern. Compare \#hi $+\mathrm{s}+\mathrm{t}+\mathrm{ater}+$ with \#hi\#s\#t\#ater\#. The domains enclosed by the root boundary marker ' + ' are $+\mathrm{s}+,+\mathrm{t}+$, and + ater + .

[^40]:    ${ }^{24}$ Since Semelai (Austroasiatic, Malaysia) and Yeri (Nuclear Toricelli, Papua New Guinea) have two differing infixes, this yields 55 interaction patterns.

[^41]:    ${ }^{25}$ I use $\left[\mathrm{s}^{*}\right]$ as an IPA-compatible notation for what Durie (1985:12) calls a laminal alveodental fricative with a wide channel area. An alternative notation might be [s]. Durie (1985) notates $[\mathrm{S}]$.

[^42]:    ${ }^{26}$ Some languages have additional constraints on monomorphemic domains that do not interact with infixation．This accounts for the patterns in Austronesian languages with no interaction．

[^43]:    ${ }^{27}$ Of course，this requires some additional mechanism for paradigm gaps in Optimality Theory，cf．e．g． Orgun \＆Sprouse（1999）．

[^44]:    ${ }^{28}$ Reiner (1965) groups these consonants as dentals and sibilants, whereas Lubowicz (2010) merges them into a group called coronals. The set includes $/ \mathrm{t} \mathrm{d} \mathrm{t}^{\prime} /$ and $/ \mathrm{s}, \mathrm{z}, \mathrm{ts}$ '/ in both description and excludes $/ \mathrm{f} /$ and all non-coronals, i.e. labials, palatal, velar and post-velar consonants and all sonorants. The difference might be due to the fact that only Reiner (1965) distinguishes between reversible and non-reversible orderings.
    ${ }^{29}$ I changed the transcription to IPA based on Huehnergard (1997): $s{ }_{\mathrm{s}}=\int, \mathrm{s} / \mathrm{S}=\mathrm{ts} \mathrm{s}^{\prime}, \mathrm{T} / \mathrm{t}=\mathrm{t}^{\prime}, \mathrm{q}=\mathrm{k}{ }^{\prime}$.
    ${ }^{30}$ Note that Akkadian has several processes that alter syllable structure, e.g. pervasive syncope. For reasons of space, these will not be described here in further detail.

[^45]:    ${ }^{1}$ This term is not mentioned in Clements $(1984,1985)$. Alternatively, one could talk about restrictions on vowel harmony triggers (cf. van der Hulst 2018), but such a term has a more general meaning and includes directionality (trigger has to appear at an edge) and stress-triggered vowel harmony.
    ${ }^{2}$ Clements (1985) transcribes this vowel as [a], whereas Clements (1984) uses the IPA symbol [ə]. I will use the latter in this chapter.

[^46]:    ${ }^{3}$ It should be noted that the low vowels can sometimes act as opaque vowels in that they create disharmonic words, e.g. [funãã] 'to rekindle' (Joana Serwaa Ampofo, p.c.) and [kaŋkəbi] 'millipede' (Clements 1985:62). This also includes [+ATR] low vowels [ə] that fail to trigger harmony, as in [ 3 -kəri-i] 'he weighed (it)' (Clements 1985:63). Clements (1984) does not actually include /a/ in the general vowel harmony rule but postulates a specific late raising rule, see section 4.4.1.
    ${ }^{4}$ Here and in the following, marks an optimal but unattested output and $\circledast$ marks an attested output that does not become optimal.

[^47]:    ${ }^{5}$ The main idea of the present approach is also similar to LiCEnSE-based approaches (Walker 2005, 2010, 2011; Kaplan 2011, 2018, 2019) in that vowel harmony is construed of as an asymmetric relation. Iacoponi (2018) also proposes a similar analysis for a different kind of pattern in Basque sibilant harmony.
    ${ }^{6}$ McCarthy (2004) himself only uses constraints on heads to derive directionality and actually uses the inverse kind of constraints by requiring certain segments to be the head of their span in order to derive blocking effects in nasal harmony.

[^48]:    ${ }^{7}$ Finley (2008:43) notes that splitting the harmony driver from the directionality constraints makes pathological predictions with regard to a specific version of the sour grapes problem. If directionality constraints are ranked low, harmony can be birectional. However, if a blocker is present, directionality emerges. Finley proposes a solution where the two constraints are combined into a single Spread constraint. Nothing hinges on either option in the following analysis, since directionality constraints and trigger constraints are never crucially ranked with regard to each other in the case studies presented here.

[^49]:    ${ }^{8}$ A third candidate $[(\underline{p})(\underline{y})(\underline{a w a})(\underline{\text { sat }})(\underline{n})]$ cannot become optimal because it incurs more violations of the harmony-driver constraint AdjacentSpan(Nasal).

[^50]:    ${ }^{9}$ Here and in the following, I assume that the constraint requiring faithfulness to a [ $\pm$ high] value is always undominated and thus changes from mid vowels to high vowels are not a possible repair strategy.
    ${ }^{10}$ Even if the cover constraint is split up, this does not help. Ranking a directionality constraint lower than ${ }^{*} e, o$, yields directionality reversal for a form like $/ \mathrm{pet} /-/ \mathrm{u} /[\mathrm{petu}]$ where $[\mathrm{e}, \mathrm{o}]$ would occur as the result of vowel harmony. Such a ranking would predict the unattested output [petv].

[^51]:    ${ }^{11} \mathrm{~A}$ derived environment pattern here refers to cases, where a certain change is only allowed if another feature change has also applied. This includes a specific pattern, where a certain segment on a scale is skipped. This specific pattern is also known as a saltation (Hayes \& White 2015; Smith 2018).

[^52]:    ${ }^{12}$ I have not attempted to give a headed spans analysis of this pattern in Akan, since the entirety of the complex empirical data seem to still be a matter of debate. Vowel harmony seems to affect low vowels only for some speakers (cf. Clements 1985:91) and has been argued to be phonetic for other speakers. This is contested by a thorough analysis of experimental data in Casali (2012). The domain of vowel harmony does not align with the prosodic word (cf. Kügler 2015) and the low vowel is opaque in some cases but not in others. See also Ampofo \& Rasin (2021) for a recent discussion. Assamese and Päri offer more clear-cut examples of an asymmetric non-trigger pattern.

[^53]:    ${ }^{13}$ Mahanta (2008) uses the constraint * $[-\mathrm{ATR}][+\mathrm{ATR}]$ as a harmony driver, which creates additional problems but does not make a difference here, since we are only concerned with faithfulness constraints, cf. Tebay \& Zimmermann (2020) for a discussion. It is worth noting that it still partially encodes directionality.

[^54]:    ${ }^{14}$ Trivially, /o/ would also map to [o], since the faithful candidate would violate neither the markedness nor the faithfulness constraints.

[^55]:    ${ }^{15}$ The constraint would make the same kind of prediction if the first part was violated for output occurrences of high vowels. If the following ranking were employed: * + high $_{\text {Output }} \& \operatorname{ID}(\mathrm{ATR}) \gg *[-\mathrm{ATR}]$ $\gg$ [-high], the following chain shift would apply: $\lrcorner \rightarrow 0, o \rightarrow u$.
    ${ }^{16}$ Again, input /o/ would trivially map to [o] since it satisfies all faithfulness and markedness constraints.

[^56]:    ${ }^{17}$ van der Hulst uses vowel harmony in a broader sense to also include cases of Umlaut, Metaphony and local non-iterative vowel assimilations. These fall within the scope of Headed Spans Theory if we allow restrictions on the size of a span to derive bounded or non-iterative spreading.

[^57]:    ${ }^{18} \mathrm{~A}$ further romance language with asymmetric non-triggers $[\mathrm{e}, \mathrm{o}]$ is Campidanese Sardinian (IndoEuropean, Italy), as shown in Torres-Tamarit et al. (2017).
    ${ }^{19}$ A further analogous pattern is reported in Eastern Mari (Uralic; Kazakhstan, Russia) where [æ] is banned from triggering rightwards vowel harmony in the initial syllable (Vaysman 2009:89).

[^58]:    ${ }^{1}$ See Inkelas \& Shih (2015) and Shih \& Inkelas (2018) for a similar argument and McPherson \& Dryer (2021) for a possible counterexample from Poko-Rawo (Skou; Papua New Guinea) based on floating tones.

[^59]:    ${ }^{2}$ There is a problem lurking here that is sometimes glossed over in prosodic phonology: resyllabification. One possible solution would be to allow for multidominance structure and assign the onset to the left mother by convention. This problem does not concern us further, since most tonal RotB problems are not affected by resyllabification. See also chapter 6 for a general discussion.

[^60]:    ${ }^{3}$ This is similar to Tesar (2013)'s concept of relative similarity. The definition here is specific to correspondence theoretic OT, whereas Tesar (2013)'s is not.

[^61]:    ${ }^{4}$ Chungli Ao also has prefixes that will be excluded from the present analysis.
    ${ }^{5}$ Temsunungsang (2009) also includes the conative suffix -tāŋ in this group, which always overwrites the root tone with a mid tone. I discuss it with other overwriting suffixes.

[^62]:    ${ }^{6}$ Cells marked with a dagger $\dagger$ indicate that Temsunungsang (2009) describes the tone pattern here without presenting concrete data. Added forms are taken from Temsunungsang (2017).

[^63]:    ${ }^{7}$ I assume that higher affixes can be morphologically specified with a requirement to be parsed into a non-minimal prosodic word. They subcategorize for their prosodic attachment, cf. e.g. Kalin \& Rolle (2022).
    ${ }^{8}$ A further candidate could shift the mid tone to the root, overwrite its tones and insert a default tone for the affix. I assume that such a candidate would lose due to faithfulness considerations.

[^64]:    ${ }^{9}$ Note that this $\mathrm{OCP}(\mathrm{H})$ constraint only refers to associated tones and therefore does not restrict the melodies at the tonal tier per se.

[^65]:    ${ }^{10}$ Note that this does not interact with the minimality requirement, since Temsunungsang (2009) explicitly argues that the minimality requirement for verbs makes reference to syllables and not to feet, since some footing is weight-sensitive but the minimality requirement on minimal prosodic words is not.

[^66]:    ${ }^{11}$ Note that these constraint also restrict monosyllabic feet to mid and high tones - a welcome result, since affixes only ever show this configuration. Monosyllabic feet are not possible for newly created feet since the foot binarity constraint is ranked below the faithfulness constraint on tones and thus only applies to epenthetic feet.
    ${ }^{12}$ See chapter 4 for an introduction and motivation of a headed version of autosegmental spans.
    ${ }^{13}$ There is an additional problem arising in cases where both the root and the affix are footed lexically with an equally marked foot. The affix foot needs to win out in these situations. I do not know of a simple constraint in Correspondence Theory that would yield this result. I leave this question open for future research.

[^67]:    ${ }^{14}$ For concreteness sake, I assume the ranking $\operatorname{IdEnt}(\mathrm{T}) \gg \mathrm{DEP}(\mathrm{T}) \gg \operatorname{MAx}(\mathrm{T}) \gg * \mathrm{H} \gg * \mathrm{M}$

[^68]:    ${ }^{15}$ Note that the resyllabification problem shows up here again. Checked syllables before trigger suffixes should resyllabify. There are two possible solutions to this problem. Either, resyllabification is blocked with trigger suffixes or a multidominance structure is created, cf. footnote 2 .

[^69]:    ${ }^{16}$ Compare also Optimal Tone Mapping as proposed in Zoll (2003), which assumes positional markedness instead of directional tone mapping but still tentatively posits underlying tone inventories.
    ${ }^{17}$ Further arguments against tonal inventories in Mende and an analysis based on Agreement by Correspondence and Q-Theory are provided in Inkelas \& Shih (2015).

[^70]:    ${ }^{18}$ Here and in the following, I assume that - in the absence of evidence to the contrary - feet occur in at the edges of a prosodic word and can be diagnosed by the segmental and supreasegmental phonological patterns of a given language, including the distribution of tone.
    ${ }^{19}$ Note that this constraint only applies to associated tones. An $\mathrm{OCP}(\mathrm{H})$ constraint that solely refers to the tonal tier is not needed.

[^71]:    ${ }^{20}$ This can be seen as a cover constraint that penalizes both one tone linked to two moras and one mora linked to two tones as well as floating tones and unspecified moras. It will be left out of the tableaux, since I assume that it is never violated in Mee and therefore undominated.

[^72]:    ${ }^{21}$ Alternatively, this constraint could require the final foot to be linked to at least one high tone. This would require an additional $\mathrm{OCP}(\mathrm{H})$ constraint, as employed in the analysis of Mende.

[^73]:    ${ }^{22}$ This section heavily draws from the arguments against directional tone mapping in Zoll (2003).

[^74]:    ${ }^{23}$ A similar analysis is proposed for tone in Kukuya (Atlantic-Congo; Congo, Gabon) in Archangeli \& Pulleyblank (1994:345). Topintzi (2003) has proposed an analysis of K. Shona based on high tone attraction to foot heads, cf. also Breteler (2017) for a similar approach, and Jardine (2016) for criticism.
    ${ }^{24}$ This would be a universal restriction on underlying representations and thus be compatible with Richness of the Base, which only bans language-specific constraints.

[^75]:    ${ }^{25}$ For exactly this reason, most of the languages discussed as classical examples of early Autosegmental do not show up in DoCoMD.

[^76]:    ${ }^{26}$ Note that I conflated root and root proper domains here and ignored category-specificity. See the next subsection for a discussion of the latter.

[^77]:    ${ }^{27}$ Even if it were restricted to noun roots, this would only slightly complicate the analysis. An additional constraint on non-minimal noun words would have to introduced.

[^78]:    ${ }^{28}$ The list includes Khayo (Kenya, Uganda), Marachi (Kenya), Nyala-West (Kenya), Saamia (Kenya, Uganda), Tura (Kenya), Kuria (Kenya, Tanzania), Makonde (Mozambique, Tanzania), Matuumbi (Tanzania), Rufiji (Tanzania), Mwera (Tanzania), Yao (Malawi, Mozambique Tanzania, Zambia), Bena (Tanzania), Hehe (Tanzania), Kinga (Malawi, Tanzania), Safwa (Tanzania), the Ngoni dialect of Nsenga (Zambia), Kanyok (Democratic Republic of Congo), the Ipila dialect of Kete (Democratic Republic of Congo), and Ruwund (Angola, Democratic Republic of Congo).

[^79]:    ${ }^{1}$ Perhaps surprisingly, it is not clear if Stratal OT violates Output Drivenness in its narrowest sense, since the mapping at each individual stratum can be assumed to be output-driven. The composition of

[^80]:    several output-driven mappings does not need to be output-driven, cf. Tesar (2013:170).
    ${ }^{2}$ Note that the possibility of root domain constraints to apply late is independently needed to derive the Muna type of interaction where infixes become part of the root domain for root domain constraints and undergo repair mechanisms in order to satisfy it.

[^81]:    ${ }^{3}$ Note however, that this does not necessarily mean that Constraints on Underlying Representations pose a problem for learnability. As shown by Rasin \& Katzir (2020) Constraints on Underlying Representation can ease phonological learning under certain assumptions. A parallel argument for complex faithfulness

[^82]:    constraints has not been made in the literature.
    ${ }^{4}$ Again, late root domain constraints are needed to derive the fact that infixes undergo repair mechanisms in order to satisfy root domain constraints in Muna.

