



**HAL**  
open science

# The phonological implications of the metrical properties of the faithfulness constraints

Giorgio Magri

► **To cite this version:**

Giorgio Magri. The phonological implications of the metrical properties of the faithfulness constraints. Linguistics. Université Paris 8, 2017. tel-01691437

**HAL Id: tel-01691437**

**<https://hal.science/tel-01691437>**

Submitted on 14 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris 8  
UFR Sciences du Langage  
Habilitation a diriger des recherches  
Paris, January 2017

**Giorgio Magri**

**The phonological implications of the metrical properties of the  
faithfulness constraints**

**Tome 1: Synthèse & perspectives**

Soutenue le 1er mars 2017

**Membres du jury:**

Paul Boersma (University of Amsterdam)  
Francois Dell (CNRS)  
Emmanuel Dupoux (EHESS)  
René Kager (Utrecht University)  
Sophie Wauquier (Université Paris 8; garant)



# Contents

Preface	9
Chapter 1. Research trajectory	11
1.1. Overview	11
1.1.1. Work on natural language semantics	11
1.1.2. Work on computational constraint-based phonology	11
1.1.3. Impact	12
1.2. Main results on nominal semantics and the ontology of quantification	13
1.2.1. First result: the syntax of partitives	13
1.2.2. Second result: number agreement and distributivity	14
1.2.3. Third result: mass and collective nouns	15
1.3. Main results on individual-level predicates and scalar implicatures	16
1.3.1. First result: ILPs as homogeneous predicates	16
1.3.2. Second result: oddness and scalar implicatures	18
1.3.3. Third result: oddness in DE contexts	19
1.3.4. Fourth result: ILPs as a theorem of the theory of oddness	21
1.4. Main results on error-driven learning and the mode of constraint interaction	21
1.4.1. First result: a complete computational theory of OT error-driven learning	22
1.4.2. Second result: comparison with HG error-driven learning	23
1.4.3. Third result: algorithmic portability from HG into OT	24
1.5. Main results on restrictiveness and the early acquisition of phonotactics	25
1.5.1. First result: constraint promotion is needed from a modeling perspective	25
1.5.2. Second result: intractability of the problem of the acquisition of phonotactics	26
1.5.3. Third result: restrictiveness through assumptions on phonotactic patterns	26
1.5.4. Fourth result: restrictiveness through assumptions on constraint sets	27
1.5.5. Fifth result: modeling child epenthesis data from the INANP database	28
Chapter 2. Project overview	29
2.1. Idempotency and counter-feeding opacity (section 3.1)	30
2.2. The faithfulness idempotency condition (section 3.2)	31
2.3. Which faithfulness constraints satisfy the FIC? (sections 3.3-3.5)	32
2.4. The faithfulness triangle inequality (section 4.1)	32
2.5. Idempotency and the triangle inequality: the HG case (section 4.2)	33
2.6. Faithfulness additivity, categoricity, and monotonicity (section 4.3)	33
2.7. Idempotency and the triangle inequality: the OT case (section 4.4)	34
2.8. Output-drivenness and counter-bleeding opacity (section 5.1)	35
2.9. Output-drivenness in OT (section 5.2)	35
2.10. An axiomatic definition of the similarity order (sections 5.3-5.4)	35
2.11. Output-drivenness and the triangle inequality (section 5.5)	36
2.12. Output-drivenness in HG (section 5.6)	36
2.13. A network of relationships among faithfulness conditions (section 6.1)	37
2.14. Opacity between output-drivenness and input-drivenness (sections 6.2 and 6.3)	37
2.15. Idempotency and chain shifts (section 6.4)	38
2.16. Benign chain shifts (section 6.5)	38
Chapter 3. The theory of idempotency in Optimality Theory	39
3.1. Idempotency	39

3.1.1.	Representational framework	39
3.1.2.	Identity candidates	40
3.1.3.	Idempotency	40
3.1.4.	Chain shifts	41
3.2.	Idempotency in Optimality Theory	42
3.2.1.	Classical Optimality Theory (OT)	42
3.2.2.	A sufficient condition for chain shifts	43
3.2.3.	The faithfulness idempotency condition (FIC)	44
3.2.4.	Composition candidates and the $\text{FIC}_{\text{comp}}$	44
3.2.5.	Refinements	45
3.3.	Establishing the $\text{FIC}_{\text{comp}}$ : basic constraints	45
3.3.1.	MAX	46
3.3.2.	DEP	46
3.3.3.	IDENT	47
3.4.	Establishing the $\text{FIC}_{\text{comp}}$ : restricted constraints	49
3.4.1.	$\text{MAX}_R$	49
3.4.2.	$\text{DEP}^S$	50
3.4.3.	$\text{MAX}_R^S$	50
3.4.4.	$\text{DEP}_R^S$	51
3.4.5.	$\text{IDENT}_{\varphi,R}$	51
3.4.6.	$\text{IDENT}_{\varphi}^S$	52
3.4.7.	$\text{IDENT}_{\varphi,R}^S$	52
3.5.	Establishing the $\text{FIC}_{\text{comp}}$ : other constraints	53
3.5.1.	$\text{MAX}_{[+\varphi]}$ , $\text{DEP}_{[+\varphi]}$	53
3.5.2.	UNIFORMITY and INTEGRITY	54
3.5.3.	CONTIGUITY	54
3.5.4.	ADJACENCY	55
3.5.5.	LINEARITY	55
3.5.6.	Constraint conjunction and disjunction	56
3.6.	Conclusions	57
A.	Proofs	58
A.1.	Proof of proposition 2	58
A.2.	Proof of proposition 3	59
A.3.	Proof of proposition 4	60
Chapter 4.	Idempotency and the triangle inequality	63
4.1.	The metric nature of the faithfulness constraints	63
4.1.1.	Intuition: the metric triangle inequality	63
4.1.2.	Formalization: the $\text{FTI}_{\text{comp}}$	64
4.1.3.	Faithfulness constraints and other metrical axioms	64
4.1.4.	Summary	64
4.2.	Idempotency in Harmonic Grammar and the triangle inequality	64
4.2.1.	Harmonic Grammar	64
4.2.2.	A condition on the candidate set: the one-to-one assumption	65
4.2.3.	A condition on the faithfulness constraint set: the $\text{FIC}_{\text{comp}}^{\text{HG}}$	66
4.2.4.	The $\text{FIC}_{\text{comp}}^{\text{HG}}$ is stronger than the $\text{FIC}_{\text{comp}}^{\text{OT}}$ in the general case	66
4.2.5.	The $\text{FIC}_{\text{comp}}^{\text{HG}}$ is equivalent to the $\text{FTI}_{\text{comp}}$ in the general case	67
4.2.6.	Summary	67
4.3.	Faithfulness categoricity, additivity and monotonicity	67
4.3.1.	Additivity, categoricity, and monotonicity w.r.t. correspondence relations	67
4.3.2.	Additivity, categoricity, and monotonicity w.r.t. underlying strings	69
4.3.3.	A remark on the proper definition of I-additivity	70
4.3.4.	Additivity, categoricity, and monotonicity w.r.t. surface strings	70
4.3.5.	McCarthy's (strengthened) categoricity conjecture	71
4.3.6.	Summary	72
4.4.	OT Idempotency and the triangle inequality	72

4.4.1.	The $\text{FTI}_{\text{comp}}$ is stronger than the $\text{FIC}_{\text{comp}}^{\text{OT}}$ in the general case	72
4.4.2.	The $\text{FTI}_{\text{comp}}$ and the $\text{FIC}_{\text{comp}}^{\text{OT}}$ are equivalent for binary constraints	73
4.4.3.	The $\text{FTI}_{\text{comp}}$ and the $\text{FIC}_{\text{comp}}^{\text{OT}}$ are equivalent for categorical constraints	73
4.4.4.	On the candidate assumptions of proposition 15	74
4.4.5.	On the constraint assumptions of proposition 15	74
4.4.6.	Which faithfulness constraints satisfy the $\text{FTI}_{\text{comp}}$	75
4.4.7.	The $\text{FIC}_{\text{comp}}^{\text{HG}}$ and the $\text{FIC}_{\text{comp}}^{\text{OT}}$ are equivalent for categorical constraints	75
4.4.8.	Summary	76
A.	Proofs	76
A.1.	Proof of proposition 13	76
A.2.	Proof of proposition 15	78
Chapter 5.	The theory of output-drivenness from the perspective of the triangle inequality	81
5.1.	Output-drivenness and its relationship with idempotency	81
5.1.1.	Tesar's notion of output-drivenness	81
5.1.2.	Output-drivenness entails idempotency	82
5.1.3.	Output-drivenness is stronger than idempotency: saltations	82
5.2.	Output-drivenness in OT	83
5.2.1.	A condition on the candidate set: the one-to-one assumption	83
5.2.2.	A condition on the faithfulness constraint set: the $\text{FODC}^{\text{OT}}$	83
5.2.3.	The $\text{FODC}^{\text{OT}}$ entails the $\text{FIC}^{\text{OT}}$	84
5.2.4.	The $\text{FODC}^{\text{OT}}$ is stronger than the $\text{FIC}^{\text{OT}}$	84
5.2.5.	Only a sufficient constraint condition?	85
5.3.	Tesar's concrete definition of the similarity order	85
5.3.1.	An initial attempt	85
5.3.2.	Towards a stronger similarity order	86
5.3.3.	Tesar's similarity order	87
5.4.	Axiomatizing the similarity order	87
5.4.1.	An initial attempt	88
5.4.2.	A stronger axiomatization	88
5.4.3.	Formal properties of the strong similarity order	90
5.4.4.	The axiomatic definition subsumes Tesar's concrete definition	90
5.4.5.	The composition candidate and the $\text{FODC}_{\text{comp}}^{\text{OT}}$	90
5.5.	OT output-drivenness and the triangle inequality	91
5.5.1.	The $\text{FTI}_{\text{comp}}$ entails the $\text{FODC}_{\text{comp}}^{\text{OT}}$ in the general case	91
5.5.2.	The $\text{FTI}_{\text{comp}}$ is stronger than the $\text{FODC}_{\text{comp}}^{\text{OT}}$ in the general case	92
5.5.3.	The $\text{FTI}_{\text{comp}}$ , $\text{FODC}_{\text{comp}}^{\text{OT}}$ , and $\text{FIC}_{\text{comp}}^{\text{OT}}$ are equivalent for categorical constraints	92
5.5.4.	The restriction to faithfulness constraints which belong to the subset $\mathcal{F}$	93
5.5.5.	Summary	93
5.6.	Output-drivenness in Harmonic Grammar	93
5.6.1.	A condition on the faithfulness constraint set: the $\text{FODC}_{\text{comp}}^{\text{HG}}$	94
5.6.2.	The $\text{FODC}^{\text{HG}}$ entails the $\text{FIC}^{\text{HG}}$ in the general case	94
5.6.3.	The $\text{FODC}^{\text{HG}}$ entails the $\text{FODC}^{\text{OT}}$ in the general case	95
5.6.4.	The $\text{FODC}_{\text{comp}}^{\text{HG}}$ , $\text{FIC}^{\text{HG}}$ , and $\text{FTI}_{\text{comp}}$ are equivalent in the general case	95
5.6.5.	The $\text{FODC}_{\text{comp}}^{\text{HG}}$ , $\text{FODC}_{\text{comp}}^{\text{OT}}$ and $\text{FIC}_{\text{comp}}^{\text{OT}}$ are equivalent for categorical constraints	96
5.6.6.	Summary	96
A.	Proofs	96
A.1.	Proof of proposition 18	96
A.2.	Proof of proposition 19	97
A.3.	Proof of proposition 23	99
A.4.	Proof of proposition 24	100
Chapter 6.	Implications for phonology and its acquisition	103
6.1.	Summary: idempotency, output-drivenness, and the triangle inequality	103
6.1.1.	Implications for the interpretation of faithfulness constraint conditions	104
6.1.2.	Implications for the relationship between idempotency and output-drivenness	104

6.1.3.	Implications for the relationship between OT and HG	104
6.1.4.	Implications for the analysis of faithfulness constraints	104
6.2.	Output-drivenness and opacity	105
6.2.1.	Counter-feeding and chain shifts	105
6.2.2.	Counter-bleeding and saltations	106
6.2.3.	Output-drivenness, counter-feeding, and counter-bleeding	106
6.2.4.	Refinements	108
6.3.	Output-drivenness or input-drivenness?	109
6.3.1.	Input-drivenness	109
6.3.2.	Input-drivenness entails idempotency	110
6.3.3.	Input-drivenness and saltations	110
6.3.4.	Sufficient conditions for input-drivenness	111
6.3.5.	The triangle inequality entails input-drivenness	111
6.3.6.	How to pull apart output-drivenness and input-drivenness	112
6.4.	Idempotency and chain shifts	115
6.4.1.	Only a sufficient condition?	115
6.4.2.	Chain-shifts through constraint conjunction	116
6.4.3.	Chain shifts through constraint restrictions	117
6.4.4.	Chain-shifts through breaking	120
6.4.5.	Summary	120
6.5.	Idempotency and the early acquisition of phonotactics	120
6.5.1.	Learning without underlying forms	121
6.5.2.	Reconstructing faithful underlying forms	122
6.5.3.	Idempotency	122
6.5.4.	Benign chain shifts	123
6.5.5.	Benign chain shifts derived through two markedness constraints	124
6.5.6.	Benign chain shifts derived through constraint conjunction	125
6.5.7.	Benign chain shifts derived through breaking	126
6.5.8.	Benign chain shifts derived through constraint restrictions	127
6.5.9.	Summary	128
A.	Proofs	128
A.1.	Proof of proposition 27	128
A.2.	Proof of proposition 28	129
	Bibliography	131
	Appendix A. List of publications (as of June 2016)	139
A.1.	Papers in journals	139
A.2.	Books, book chapters, and book reviews	139
A.3.	Papers in proceedings of peer reviewed conferences	140
A.4.	Manuscripts (available at <a href="https://sites.google.com/site/magrigr/">https://sites.google.com/site/magrigr/</a> )	142
A.5.	Manuscripts in preparation	142
	Appendix B. Curriculum vitae (as of June 2016)	143
B.1.	Contact information	143
B.2.	Employment	143
B.3.	Education	143
B.4.	PhD dissertation	143
B.5.	Grants	143
B.6.	Papers in journals	144
B.7.	Books, book chapters, and book reviews	144
B.8.	Manuscripts (available at <a href="https://sites.google.com/site/magrigr/">https://sites.google.com/site/magrigr/</a> )	145
B.9.	Papers in proceedings of peer reviewed conferences	145
B.10.	Talks and posters in peer reviewed conferences	147
B.11.	Invited talks at conferences and workshops	149
B.12.	Other invited talks	150
B.13.	Contributed talks	150

B.14. Teaching	150
B.15. Conference organization	151
B.16. Academic service	152





## Preface

My research over the last ten years has contributed to two different and unrelated subfields of linguistic theory: natural language semantics and pragmatics; and computational phonology from a constraint-based perspective. Despite spanning two unrelated sub-fields, my research contributions are unified by a common focus on *computation*, on the *formal aspects* of the linguistic theory, on the *mathematical nature* of the underlying structure. Typological facts, linguistic evidence, and psycholinguistic data are reduced to their core formal structure, accounted through detailed derivations from a few abstract principles, subserved to the purpose of a mathematical argument. This document illustrates this property of my research. Chapter 1 provides a broad outline of my research trajectory, by summarizing the main results I have obtained so far. The rest of the volume illustrates the interplay between linguistic theory and computation which characterizes my research by outlining a project on the computational theory of opacity in constraint-based phonology. Chapter 2 provides an informal overview of the theory developed in chapters 3-6. Two final appendices provide my list of publications and curriculum vitae (as of June 2016). This volume comes with a companion volume which reproduces my journal papers. All my publications are available at <https://sites.google.com/site/magrigrg/>.

**Editorial notice:** The material presented in chapter 3 is based on Magri (to appeara). The material presented in chapters 4 and 5 is based on Magri (to appearb), Magri (to appearc), and Magri (2016b). Chapter 6 is more speculative and reports still unpublished material. Parts of this document have been presented at WCCFL 33 at Simon Fraser University in March 2015 (see also Magri 2016a); at the EPG (Experimental Phonology Group) meeting at Utrecht University in June 2015; at the LSA Workshop on Computational Phonology and Morphology at the University of Chicago in July 2015; at the Rutgers Optimality Research Group in September 2015; at OCP 12 in Budapest in January 2016; and at MIT in April 2016. Feedback from those audiences is gratefully acknowledged. Finally, I would like to thank Bruce Tesar for very useful comments on an earlier version of chapter 5. The research reported in this document has been supported in part by a Marie Curie Intra European Fellowship (Grant agreement number: PIEF-GA-2011-301938).



## Research trajectory

Section 1.1 singles out the four main topics I have worked on so far: nominal semantics and the ontology of quantification; individual-level predicates and the theory of scalar implicatures; the computational theory of error-driven learning in constraint-based phonology and the mode of constraint interaction; restrictiveness and the so called early stage of the acquisition of phonotactics. Sections 1.2-1.5 summarize the main results I have obtained so far, organized around these four main research topics. Throughout this chapter, my publications are referred to with the numerical labels assigned to them in the list of publications provided in Appendix A.

### 1.1. Overview

This section provides a broad overview of my research trajectory from graduate school to my current research, singling out four main research topics. Sections 1.2-1.5 will then outline my main contributions to each of these four topics.

#### 1.1.1. Work on natural language semantics

During my undergraduate years at the University of Milano, I have pursued in parallel the degrees in philosophy and mathematics (4 years each), which I have both completed *summa cum laude*. I have then pursued training in generative linguistics through the linguistics graduate program at MIT. My research in the first years of graduate school focused on **nominal semantics and the ontology of quantification**. I have developed a new approach to the semantics of plurals and I have explored its implications for issues such as the proper characterization of the mass/count distinction, the syntax/semantics of partitives, and the semantics of collective nouns. This research has been consolidated into [pub 3], [pub 12], [pub 14], [pub 30], and [man 4]. Section 1.2 summarizes the main results obtained.

Shortly after the beginning of graduate school, my research interests in natural language semantics have extended beyond the nominal domain. In the first part of my dissertation, I have developed a new theory of a well studied class of statives, called **individual-level predicates**. My proposal is purely pragmatic and thus dispenses with the many *ad hoc* grammatical characterizations of this class of predicates that have been proposed in the literature. Furthermore, my proposal has far reaching implications for the recent debate on **the nature of scalar implicatures**, in particular for the issues of implicature cancellation and of the modularity of implicature computation. This research has been consolidated into [pub 24], [pub 28], [pub 29], [pub 37], and [sub 1], together with various papers in conference proceedings. Section 1.3 summarizes the main results obtained.

#### 1.1.2. Work on computational constraint-based phonology

Towards the end of my graduate years, my research interests have extended to computational models of the acquisition of sound patterns. In order to gain some research experience in such a different research field, I have applied for a fifth year of graduate funding. Building on my strong undergraduate mathematical education, I have been able to take advanced classes in statistics and machine learning (at the Laboratory for Information and Decision Theory of MIT; at the Statistics and Information Sciences Laboratory of Harvard; at the Laboratory for Information and Decision Theory of MIT). During my postdoctoral years in Paris, I have vigorously pursued training in Machine Learning: I have attended machine learning graduate seminars (including Stéphane Boucheron's graduate seminar on concentration inequalities for machine learning in spring 2011 at the University of Paris 7); I have attended the meetings of the SMILE (Statistical Machine Learning) group in Paris as well as the Machine Learning Workshop organized by the French Mathematical Society in 2011; I have been admitted to the 18th Machine Learning Summer School. This solid formal training has allowed me to obtain new results on the computational theory of **error-driven learning within constraint-based phonology**, focusing in particular on the

implications of different modes of constraint interaction for issues such as convergence, efficiency, and noise robustness. This research has been consolidated into [pub 31], [pub 32], [pub 34], [pub 38], [pub 39], and [pub 40], together with various papers in conference proceedings. Section 1.4 summarizes the main results obtained.

This initial work in computational phonology built on my education in mathematics and machine learning and focused on the purely computational perspective of language learnability. Towards the end of my postdoctoral years, I have started to complement this learnability perspective with rich child acquisition data. A grant from the Fyssen Research Foundation in 2011 allowed me to pursue this research at the interface of mathematics and acquisition in collaboration with Adam Albright from MIT. I have focused on the so called **early stage of the child’s acquisition of phonotactics**. How does the error-driven learning scheme investigated since my dissertation work fare in modeling this early acquisition stage? I have tackled this question from both a learnability and an acquisition perspectives. From the learnability perspective, I have developed analytical guarantees that the model succeeds at learning a sufficiently restrictive phonotactics. From an acquisition perspective, I have explored (in collaboration with Albright) the model’s predictions on epenthesis data extracted from the INANP (*Iowa-Nebraska Articulation Norms Project*) database (Smit et al. 1990), that I have acquired from Ann Bosma Smit at Kansas State University. This research has been consolidated into [pub 16], [pub 18], [pub 19], [pub 20], [pub 23], [pub 33] and [pub 36]; this research will be further extended through [in prep 1], [in prep 2], [in prep 3], and [in prep 4]. Section 1.5 summarizes the main results obtained.

In order for my work in computational phonology to reach maturity, it needed to be fertilized through a closer connection with core phonological theory. A Marie Curie Fellowship for Career Development has allowed me to strengthen my expertise in phonetics, typology and constraint-based phonology under the guidance of René Kager at Utrecht University in the academic years 2013-2015. This new expertise has allowed me to start a new thread in my research agenda, focused on **computational opacity** in constraint-based phonology and its learnability implications. I have focused on two opaque patterns: *chain-shifts* and *saltations* (a special type of derived environment effects), which lead to the two formal properties of **idempotency** and **output-drivenness**. I have investigated these structural properties from the perspective of sophisticated theories of phonological constraints, such as McCarthy and Prince’s (1995) *Correspondence Theory*. These results have been consolidated into [pub 21], [pub 26], [sub 2], [sub 3], [sub 4], and [man 3]. The rest of this document outlines the theory thus obtained.

### 1.1.3. Impact

My work in natural language semantics has had a strong impact on the literature. This is demonstrated not only by the sheer number of citations (my 2009 paper [pub 28] in the journal *Natural Language Semantics* has received over 115 citations according to Google Scholar), but also by the number of articles and dissertations which have engaged with my proposals on scalar implicatures and individual-level predicates. Here is a partial list. Spector (2014) is a detailed commentary of my proposal on the computation of scalar implicatures. In his MIT dissertation, Thomas (2012) provides support for (together with some modifications of) my analysis of life-time effects based on data from Mbyá Guaraní. In his Stanford dissertation, Lauer (2014) takes issue with my theory of mandatory scalar implicatures. And Romoli (2012) follows up on that issue. In her MIT dissertation, Meyer (2013) discusses my theory of blind scalar implicatures and develops an alternative. Another variant is developed in Schlenker (2012), which I actually show to be identical to my original proposal in [sub 1]. Further discussion of my work is provided in Singh (2009, 2010, 2012) and Katzir and Singh (2014). Finally, I would like to mention the fact that, when I first arrived in Paris at the beginning of my postdoctoral period in 2009, Angelika Kratzer (who was at the time on sabbatical in Paris) organized a workshop at the Maison Suger dedicated to my work, as another indication of its impact on the field.

My work in phonological learnability is more recent, but has already attracted the attention of the research community. The specific implementation of error-driven learning developed in my 2012 paper [pub 31] in *Phonology* has become standard in the OT modeling literature. And I have been recently invited by the journal *Phonology* to write the review of Bruce Tesar’s recent book, showing that I have been able to establish myself as one of the leading experts in computational phonology. Finally, I have been invited to give a number of mini-courses on computational phonology at venues such as the École Normale Supérieure, the University of Utrecht, the University College London, and MIT. I have also taught courses on computational phonology at summer schools such as ESSLLI and the LSA summer institute.

## 1.2. Main results on nominal semantics and the ontology of quantification

The lexical distinction between *singular* and *plural* nominal morphology is at the heart of the proper semantics of count, collective and mass nouns. A classical strategy to account for the linguistic facts that pertain to this distinction is to “ontologize” the distinction into a distinction between singular and plural individuals (Link 1983; Landman 1991; Chierchia 1996). The core idea is that the domain of quantification is endowed with an *atomic* part-of relation that sorts the individuals into those that have proper parts (*plural* individuals) and those that don’t (*atomic* individuals). And the morphological number distinction is then mapped onto that ontological distinction, by assuming that a morphologically singular term such as ‘the boy’ denotes an atomic individual while a morphologically plural term such as ‘the boys’ denotes a plural individual. Yet, this approach runs into troubles with singular partitives, such as ‘some of this boy’: the determiner ‘some’ seems to quantify over parts of the individual denoted by ‘this boy’, despite that individual being atomic, as it is the denotation of a singular term. To get around this problem, this literature assumes that the domain of individuals is also endowed with a further *non-atomic* part-of relation; that the atomic individual denoted by the singular term ‘this boy’ does have proper parts relative to the latter part-of relation; and that it is these parts that ‘some’ quantifies over in ‘some of this boy’.

My research on nominal semantics explores the hypothesis of a simplification of this ontology of individuals that does away with the atomic part-of relation, and thus with the ontological distinction between atomic and plural individuals that comes with it. The core intuition is that the set of atomic individuals is not encoded once and for all into the algebraic structure of the domain of quantification. Rather, it is provided whenever needed by a proper sortal fed by the syntax to the semantics in the form of a phonologically null noun phrase. Crucially, a given individual might count as atomic relative to a sortal but as non-atomic relative to another sortal. [pub 3], [pub 12], [pub 14], [pub 30], and [man 4] outline this *sortal theory of plurals* and start to explore some of its implications for the semantics of mass and count nouns. This section outlines the main results I have obtained so far.

### 1.2.1. First result: the syntax of partitives

Various authors have provided syntactic arguments for a hidden sortal in partitives above the lower definite, so that the LF of (1a) boils down to something like *some BOYS of these boys*, featuring the covert noun BOYS (Selkirk 1977 and Cardinaletti and Giusti 2006 among others).

- (1) a. Some of these boys.  
b. Some of this boy.

This syntactic literature treats a plural partitive like (1a) and a singular partitive like (1b) on a par. This makes good sense: from a syntactic perspective, the choice of the number feature on the embedded noun should not imply any difference between the LFs of (1a) and (1b). Nor would we expect any difference between the LFs of (1a) and (1b) from a semantic perspective, provided that the domain of quantification is endowed with both an atomic and a non-atomic part-of relations. In fact, thanks to the two part-of relations, the interpretation of singular and plural partitives is completely analogous. In the case of the plural partitive (1a), we first use the atomic part-of relation to construct the set of atomic parts of the individual denoted by the plural definite ‘these boys’, namely the set of boys; then, we apply the determiner ‘some’ to that set. In complete analogy, in the case of the singular partitive (1b), we first use the non-atomic part-of relation to construct the set of parts of the individual denoted by the singular definite ‘this boy’, namely the set of his body parts; then, we apply the determiner ‘some’ to that set.

Rather different predictions follow from my hypothesis of a parsimonious ontology of quantification, whereby the domain of quantification is only endowed with the non-atomic part-of relation. Again, the singular partitive (1b) can be interpreted straightforwardly by means of the non-atomic part-of relation, by feeding the determiner ‘some’ with the set of non-atomic parts of the individual denoted by the embedded singular definite. Things are very different for the case of the plural partitive (1a). We cannot feed the upstairs determiner ‘some’ with the set of non-atomic parts of the individual denoted by the definite, as this set would contain legs and arms, that we do not want ‘some’ to quantify over in the case of (1a). For the case of plural partitives it is thus crucial to have the right sortal above the definite, which provides the relevant granularity. In conclusion, my hypothesis of a parsimonious ontology of individuals predicts an otherwise unexpected asymmetry between singular and plural partitives: that the latter require a hidden sortal, while the former don’t. My first result in nominal Semantics is that this

prediction is borne out. In [pub 3], I review from the syntactic literature various arguments for a covert noun in partitives and I show that the arguments don't quite hold for the case of singular partitives.

Here is an example that illustrates the flavor of the argument. Selkirk (1977) notes an ambiguity with relative clauses in partitives: the several paintings by Sieneese artists in (2a) can be either among the paintings they saw or among the famous paintings in the museum and not necessarily seen by them; the ambiguity is lost in the case of the non-partitive (2b). Under the assumption that partitives have two copies of the noun 'paintings' (the actual one plus a copy above the downstairs determiner), this ambiguity is straightforwardly accounted for: the relative clause can modify either the higher or the lower noun.

- (2) a. In the Uffizi they saw many of the famous paintings, several of which were by Sieneese artists.  
 b. In the Uffizi they saw many famous paintings, several of which were by Sieneese artists.

Selkirk only considers plural partitives, such as (2). Let's now turn to singular partitives, by comparing the two pairs (3) and (4). Sentences (3) contain a plural partitive: the relative clause 'most of which ...' is ambiguous in the way detected by Selkirk, with no difference between (3b) with an overt sortal above the definite and (3a) without it. The case of the singular partitive in (4) is different: the relative clause is ambiguous only in the case of (4b) with the overt sortal but not in the case of (4a) without it.

- (3) a. In the library, they read some of those books, most of which were interesting.  
 b. In the library, they read some *books* of those, most of which were interesting.  
 (4) a. In the library, they read some of that book, most of which was interesting.  
 b. In the library, they read some part of that book, most of which was interesting.

The contrast between (3) and (4) suggests that there is no sortal for the relative clause to modify in the case of singular partitives, contrary to the case of plural partitives.

### 1.2.2. Second result: number agreement and distributivity

The predicate 'tall' in sentence (5a) *distributes* from the plural definite down to the singular boys, so that (5a) is truth-conditionally equivalent to (5b), with a universal quantifier.

- (5) a. The boys are tall.  
 b. Every boy is tall.

It is standard to account for this equivalence by positing a covert universal quantifier for the case of (5a), called the *distributivity operator*. The set of single boys needs to be available again for this covert operator, after it has been fed as an argument to the definite article. In a semantics endowed with the atomic part-of relation, this is trivial to achieve: the set of single boys can be reconstructed as the set of atomic parts of the individual denoted by the plural definite 'the boys'. But an alternative strategy is needed under my hypothesis of a parsimonious ontology of individuals, whereby the domain of quantification is only endowed with the non-atomic part-of relation. I want to explore the following strategy. Various authors have suggested that the subject definite in (5a) is base generated in a more embedded position and then moved to its surface position. Recent theories of syntactic movement assume that this movement operation leaves in situ a copy of the dislocated subject. And Fox (1995) argues that the copy of the subject left in situ is (roughly) stripped of its determiner and turned into a property. I suggest that it is this property that provides the relevant granularity for the distributivity operator. This link between distributivity and movement predicted by my hypothesis of a parsimonious ontology of individuals provides a rationale for a generalization by Kratzer (in progress) and Ferreira (2005), that only definites that have undergone movement (and thus have left in situ a copy) can be interpreted distributively.

An interesting test case for the proposal just sketched is provided by subjects with a rich left periphery, such as collective terms ('that group of boys') and measure phrases ('some boxes of marbles'). In these cases, Fox's theory of copies leaves it open how much of the left edge of the copy left in situ gets stripped: only little, as in '~~some~~ boxes of marbles'; or a bit more, as in '~~some boxes of~~ marbles'. My proposal that the granularity of distributivity is provided by the content of the copy, predicts that we should get distributivity to the boxes in the former case and distributivity down to the marbles in the latter case. In order to test whether this correlation holds, proper tools to access the content of the covert copy are needed. One such tool is *Antecedent Contained Deletion* (ACD), according to recent accounts

of the latter, such as Fox (2002). Another tool to access the content of the covert copy is subject-verb agreement, under the assumption that agreement is established locally, and thus holds with the copy left in situ rather than with the dislocated copy. My second main result in nominal Semantics is that this predicted correlation between agreement and distributivity is borne out by certain facts in British English (see also Elbourne 1999), as argued in [pub 12].

Here is an example, that gives the flavor of the argument. Consider the new piece of data (6) on collective nouns in British English.<sup>1</sup> These data show that plural agreement is required in the case of the inherently distributive predicate ‘to be odd’. There is thus a correlation between subject/predicate number agreement and distributivity.

- (6) a. \*This set of numbers *is* (all) odd.  
 b. This set of numbers *are* (all) odd.

I show that these data follow from my sortal theory of plurals. The core intuition of the account is as follows. Assume that number agreement is established with the copy of the subject left in situ within VP. This means that this copy is ‘This set of numbers’ in the case of singular agreement in (6a) and ‘This set of numbers’ in the case of plural agreement in (6b). As it is the meaning of the copy that provides the granularity of distributivity, it is only in the case of (6b) that the predicate ‘odd’ can distribute down to the numbers, thus explaining the contrast in acceptability displayed in (6).

### 1.2.3. Third result: mass and collective nouns

There are a variety of cases where the atomic part-of relation is standardly exploited in the literature, besides the cases of partitives and distributivity just considered. Each such case raises an interesting challenge for my hypothesis of a parsimonious ontology of quantification that dispenses with the atomic part-of relation. I am interested in two such cases in particular. One case is that of the linguistic distinction between *count* and *mass nouns*, that is usually represented by assuming that the property denoted by the former is atomic (i.e. it has minimal parts) while the property denoted by the latter is atomless (i.e. it lacks minimal parts). Another case is that of the linguistic distinction between *singular* and *plural nominal morphology*, that is usually mapped into an ontological distinction between atomic and non-atomic individuals, as noted above. Both these characterizations are lost under the hypothesis that the domain of quantification is only endowed with a non-atomic part-of relation. Thus, both of these classical characterizations need to be rethought from scratch. I submit that this is a welcome conclusion. In fact, the classical characterization of the mass/count distinction in terms of the atomic/atomless distinction just mentioned has come into question in the recent psycholinguistic literature (Gillon et al. 1999; Barner and Snedeker 2005), motivating a very different approach such as the one pursued in [man 4]. Also the classical characterization of singular/plural morphology in terms of the atomic/non-atomic distinction leads into trouble, for instance w.r.t. the issue of the semantics of *collective nouns*. The core problem is whether the collective definite (7a) and the corresponding simple definite (7b) are coreferential.

- (7) a. that pack of dogs.  
 b. those dogs.

On the one hand, the pre-theoretical intuition suggests that (7a) and (7b) are indeed coreferential. Yet, this position is untenable if the morphological distinction is mapped into an ontological distinction: the two definites cannot be coreferential, since (7a) is morphologically singular and thus denotes an atomic individual while (7b) is morphologically plural and thus denotes a plural individual. This difficulty evaporates as soon as we give up the atomic part-of relation and the ontological distinction between atomic and non-atomic individuals that comes with it.

In [pub 12], I develop this intuition into a new semantics for collective nouns. I argue that a collective term such as (7a) and the corresponding individual term (7b) indeed denote the same individual. This *coreferentiality hypothesis* is the null hypothesis, since it doesn’t require any enrichment of the ontology with new individuals, such as Link’s (1984) *impure atoms*, Landman’s (1989) *groups* or Schwarzschild’s (1996) *bunches*. The argument in favor of this coreferentiality hypothesis is twofold. On the *one hand*, I provide new evidence that the mutual replacement of a collective term with the corresponding individual term most often doesn’t affect neither the truth-conditions nor the grammaticality of a sentence, even in the context of distributive predication. Since distributive predication is peculiar to plural-denoting terms,

<sup>1</sup>The judgment reported in (6) was provided to me by Paul Elbourne (p.c.).



a collective term cannot denote an atomic individual. On the *other hand*, I show how to account under this coreferentiality hypothesis for some linguistic evidence that at first sight seems to point against it. I consider two such pieces of evidence: the special behavior of collective terms with reciprocals as well as predicate non-sharing between a collective term and the corresponding individual term.

### 1.3. Main results on individual-level predicates and scalar implicatures

Predicates such as ‘tall’ or ‘French’, which intuitively denote (quasi) permanent properties, are called *individual-level predicates* (henceforth: ILPs), after Carlson (1977). A huge number of peculiar properties of this class of predicates have been noted in the literature; see Fernald (2000) for an overview. One such property is that we cannot say (8a). My starting point is the intuition that this sentence (8a) sounds odd because it triggers the inference (more precisely: the *scalar implicature*) that John is not always tall, which mismatches with the piece of common knowledge that tallness is a permanent property.

- (8) a. #John is sometimes tall.  
b. #Some Italians come from a warm country.

This reasoning carries over to the oddness of other sentences that have nothing to do with ILPs, such as (8b). In fact, I submit that this sentence (8b) sounds odd because it triggers the scalar implicature that not all Italians come from a warm country, which mismatches with the piece of common knowledge that all Italians come from the same country.

The main result of my work in semantics has been the development of this intuition into a new implicature-based theory of ILPs in [pub 24], [pub 28], [pub 29], [pub 37], and [sub 1]. I have developed a general theory of oddness, based on scalar implicatures that mismatch with common knowledge. The theory is casted within the recent *grammatical approach* to scalar implicatures (Chierchia 2004; Fox 2007; Chierchia et al. 2012). And it is developed independently of ILPs, based on various patterns of oddness such as (8b), that have nothing to do with ILPs. I have then shown how the theory of ILPs can be derived as a theorem of this more general theory of oddness, by exploiting formal analogies such as the one in (8). I have focused on various puzzling properties of ILPs that have to do with restrictions on the interpretation of their bare plural subjects, noted in Carlson (1977), Milsark (1977), and Fox (1995); with restrictions on German word order and extraction, noted in Diesing (1992); and restrictions on Q-adverbs, noted in Kratzer (1995). The main advantage of my theory of ILPs is that it does away with *ad hoc* grammatical characterizations of ILPs, such as those by Kratzer (1995), Diesing (1992) or Chierchia (1995); and instead derives the relevant properties of ILPs from the mere assumption that they denote (quasi) permanent properties, as the intuition suggests. This section sketches the four main results I have obtained so far.

#### 1.3.1. First result: ILPs as homogeneous predicates

In order to account for the many linguistic properties of ILPs, various grammatical characterizations of ILPs have been put forward in the literature: as predicates with a special argumental structure, as in Kratzer (1995); as predicates with a special Davidsonian argument, as in Jäger (2001); as predicates with a special syntactic feature, as in Chierchia (1995); as predicates selected by a special inflectional node, as in Diesing (1992); etcetera. I want instead to explore the null hypothesis that all what is special about an ILP such as ‘tall’ is that it is *homogeneous* w.r.t. time, in the sense that it either holds of an individual at every time in his life-span or else it never holds of him. This characterization of ILPs makes the following remarkable empirical prediction: that there should be no properties peculiar to ILPs; namely, that it should be possible to mimic alleged peculiar properties of ILPs with any other predicate that is homogeneous w.r.t. one of its arguments. My first main result is that this surprising empirical prediction is indeed borne out, as argued in [pub 24] and [pub 28]. Here are a few examples to illustrate the gist of the argument.

As illustrated in (9), ILPs do not tolerate time Q-adverbs; see Kratzer (1995). Let me show that this property extends to arbitrary homogeneous predicates.

- (9) a. #John is sometimes tall.  
b. #John is always tall.

The predicate ‘to come from a warm country’ in (10) is homogeneous w.r.t. Italians, since it is either the case that Italy is warm and thus all Italians come from a warm country or else Italy is cold and thus no Italians come from a warm country.

- (10) a. #Some Italians come from a warm country.  
 b. #All Italians come from a warm country.

Sentences (10) thus show that the apparently peculiar property of ILPs of disallowing overt quantification over their homogeneous time argument extends to the homogeneous argument of arbitrary homogeneous predicates.

Let me now consider a couple of slightly more involved examples. One of the most puzzling properties of ILPs concerns restrictions on the interpretation of their bare plural subjects (henceforth: BPSs). The BPS ‘firemen’ of the non-ILP ‘available’ in (11a) can be construed both existentially (“There are firemen who are available”) and generically (“Firemen are available people”); but the BPS of the ILP ‘tall’ of sentence (11b) lacks the existential reading and can only be construed generically (Milsark 1977; Carlson 1977).

- (11) a. Firemen are available.  $\sqrt{\exists}$ -BPS = There are firemen who are available  
 b. Firemen are tall.  $^*\exists$ -BPS = \*There are firemen who are tall

The Logical Form corresponding to the unattested existential reading of (11b) would plausibly look like (12): the time argument  $t$  of the ILP ‘tall’ would be bound by a covert universal (generic) operator  $\forall_t$ ; the variable  $x$  over firemen introduced by the BPS would be bound by an existential operator  $\exists_x$ ; the latter existential operator would scope below the universal operator, since existential BPSs always select the narrowest possible scope, as it is well known since Carlson (1977).

- (12)  $[\forall_t [\exists_x [P_{hom}(x, t)]]]$ .

The predicate ‘won the swimming competition’ is homogeneous in the context considered in (13): for every given individual, either he won the swimming competition throughout the entire week or else he never won, since we know that the same guy always won. The predicate ‘won the running competition’ is not homogeneous, since a given individual might have won just once. The universal quantifier ‘every day’ in (13a) and (13b) corresponds to the universal operator  $\forall_t$  in (12); the indefinite ‘a fireman’ in (13a) and (13b) corresponds to the existential quantifier  $\exists_x$  in (12); fronting of the universal quantifier in (13a) and (13b) forces it to take wide scope over the indefinite, ensuring the scope configuration represented in (12).

- (13) *Context*: a competition lasted for five days, Monday through Friday; each day, two challenges are held: swimming and running; both John and Bill know that the same guy  $x$  won the swimming competition on all five days; nothing is known about the running competition:

	M	T	W	T	F
<i>swimming</i> :	$x$	$x$	$x$	$x$	$x$
<i>running</i> :	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$

John wants to know more about these competitions, and thus asks Bill for more information; Bill replies as follows:

- a. Every day, it was a fireman who won the running competition.  
 b. #Every day, it was a fireman who won the swimming competition.

The contrast in (13) thus shows that the apparently peculiar property of ILPs of disallowing Logical Forms like (12), extends to arbitrary homogeneous predicates.

Fox (1995) adds the surprising observation that the restriction on the existential reading of BPSs of ILPs illustrated in (11) is waived when the BPS is embedded under a universal quantifier: as expected, the BPS ‘Jewish women’ of the ILP ‘to be related to Chomsky’ in (14a) lacks the existential reading; in (14b), the definite ‘Chomsky’ has been replaced by the universal quantifier ‘every Jewish man’ and the BPS ‘Jewish women’ does admit the existential reading in (14b), provided that the universal object is assigned wide scope (“For every Jewish man *there are* Jewish women related to him”).

- (14) a. Jewish women are related to Chomsky.  $^*\exists$ -BPS = \*There are women related to Chomsky  
 b. Jewish women are related to every Jewish man.  $\exists$ -BPS = \*For every man, there are women related to him

The Logical Form corresponding to the attested existential reading of (14b) looks like (15): the time argument  $t$  of the ILP ‘related to’ is bound by a covert universal (generic) operator  $\forall_t$ ; the variable  $y$  over Jewish men is bound by a wide scope universal operator  $\forall_y$ ; the variable  $x$  over Jewish women introduced by the BPS is bound by an existential operator  $\exists_x$ ; the latter existential operator scopes below all other operators, since once more existential BPSs always take the narrowest possible scope.

$$(15) \quad [\forall_y [\forall_t [\exists_x [P_{hom}(x, y, t)]]]]].$$

The predicate ‘won’ is homogeneous in the context considered in (16): for every given individual and for every given competition, either that guy won that competition throughout the entire week or else he never won it. The universal quantifier ‘every day’ in (16b) corresponds to the universal operator  $\forall_t$  in (15); the indefinite ‘a fireman’ in (16b) corresponds to the existential quantifier  $\exists_x$  in (15); the universal quantifier ‘for every competition’ in (16b) corresponds to the universal operator  $\forall_y$  in (15); fronting of the two universal quantifiers in (16b) forces them to take wide scope over the indefinite, ensuring the scope configuration (15).

- (16) *Context*: a competition lasted for five days, Monday through Friday; each day, three challenges are held: swimming, running and jumping; both John and Bill know that the same guy  $x$  won the swimming competition on all five days, the same guy  $y$  won the running competition on all five days, and the same guy  $z$  won the jumping competition on all five days:

	M	T	W	T	F
<i>swimming</i> :	$x$	$x$	$x$	$x$	$x$
<i>running</i> :	$y$	$y$	$y$	$y$	$y$
<i>jumping</i> :	$z$	$z$	$z$	$z$	$z$

John wants to know more about these competitions, and thus asks Bill for more information; Bill replies as follows:

- a. #Every day, it was a fireman who won the running competition.
- b. Every day, for every competition, it was a fireman who won it.

The contrast in (16) thus shows that the apparently peculiar property of ILPs of disallowing Logical Forms like (12) but not like (15), extends to arbitrary homogeneous predicates.

These and various other similar cases show that apparently peculiar properties of ILPs can be mimicked with arbitrary homogeneous predicates. This empirical finding means that there is no need for a dedicated theory of ILPs based on specific grammatical assumptions on ILPs. Rather, we need a general theory of oddness, that is able to explain the oddness of sentences such as (10) or (13b) that have nothing to do with ILPs. My version of such a theory is sketched in subsections 1.3.2-1.3.3. The theory of ILPs then follows as a theorem of this more general theory of oddness, along the lines sketched in subsection 1.3.4.

### 1.3.2. Second result: oddness and scalar implicatures

It is well known that an existentially quantified sentence such as (17a) triggers the inference that the corresponding universally quantified alternative (17b) is false, namely that John did not do all of his homework. This inference is called a *scalar implicature*. A crucial property of this inference is that it is context-dependent: in the context of question (18a), (17a) triggers a strong not-all implicature; but it pretty much loses that implicature in the context of question (18b). Classical (*neo*) *Gricean* theories account for the flimsy, context-dependent nature of scalar implicatures by assuming that they are pragmatic, extra-grammatical inferences (Horn 2005; Geurts 2010).

- (17)
  - a. John did *some* of the homework.
  - b. John did *all* of the homework.
- (18)
  - a. How well did John do on the homework?
  - b. Who did some of the homework?

By analogy with (17a), I want to maintain that the existentially quantified odd sentence (8b)=(10a), repeated once more in (19a), triggers the analogous scalar implicature that the corresponding universally quantified alternative (19b) is false, namely that only some Italians come from a warm country. Of course, this implicature mismatches with the piece of common knowledge that all Italians come from the same country. I thus submit that the oddness of the sentence follows from this *mismatching scalar implicature*.

- (19) a. *#Some* Italians come from a warm country.  
 b. *All* Italians come from a warm country.

This intuition of a theory of oddness based on mismatching scalar implicatures faces two challenges. My second main result consists of a theory of (mismatching) scalar implicatures that is able to cope with these two challenges, developed in [pub 24], [pub 28], and [sub 1].

The first challenge is that the mismatching implicature needs to be blind to common knowledge. In fact, the target odd sentence (19a) and the corresponding alternative (19b) are equivalent once we factor in the piece of common knowledge that all Italians come from the same country, namely both sentences say that Italy is warm. If the algorithm that computes implicatures were sensitive to common knowledge, the contrast in (19) would be left unexplained. Yet, blindness to common knowledge is inconsistent with the (neo) Gricean assumption that scalar implicatures are pragmatic inferences, as pragmatic reasoning is well known to be heavily dependent on contextual knowledge (Heim 1991). The core idea of my solution to this puzzle is as follows. The (neo) Gricean pragmatic approach to scalar implicatures has been recently challenged for independent reasons, having to do with patterns of scalar implicatures in embedded positions (Chierchia 2004; Chierchia et al. 2012). And an alternative framework has been developed, that maintains that scalar implicatures are not pragmatic, extra-grammatical inferences. Rather, they are derived by a covert operator equivalent to ‘only’, capturing the intuition that sentence (17a) enriched with its scalar implicature is essentially equivalent to the variant in (20), with overt ‘only’ (Fox 2007).

- (20) John only did SOME of the homework.

I show that the proper semantics of overt ‘only’ needs to be construed as blind to common knowledge. This observation lends strong support to the conjecture that also the covert ‘only’ that allegedly derives scalar implicatures is blind to common knowledge. Scalar implicatures are thus *expected* to be blind to common knowledge.

The second challenge for a theory of oddness based on scalar implicatures that mismatch with common knowledge is that these mismatching implicatures need to be strong and mandatory. In fact, we want to rule out the possibility that they are cancelled, suspended or not computed at all, and thus odd sentences such as (19a) rescued from their oddness. Yet, plain non-mismatching implicatures are flimsy and context-dependent, as shown by cases of implicature cancellation illustrated in (18). The core idea of my solution to this puzzle is as follows. Assume that covert ‘only’ is always present at Logical Form. It negates every alternative in its domain. And its domain is defined as the set of alternatives that are logically stronger than the target sentence and furthermore relevant (Fox and Spector to appear; Fox and Katzir 2009). In a context such as (18b), where (17a) does not trigger the not-all implicature, the universally quantified alternative (17b) is not relevant in the context of utterance. Thus it does not belong to the domain of the mandatorily present covert ‘only’, and therefore does not get negated by it. In other words, implicature cancellation is just contextual restriction of the domain of covert ‘only’. The case of an odd sentence such as (19a) is crucially peculiar with this respect. In fact, the target odd sentence (19a) and its universally quantified alternative (19b) are contextually equivalent. Since the target sentence (19a) is relevant (because uttered) and since relevance is plausibly closed w.r.t. contextual equivalence (because it is a contextual notion), then the alternative (19b) is relevant too and must get negated by the covert ‘only’, thus deriving the mismatching implicature. Mismatching implicatures are thus *expected* to be mandatory.

### 1.3.3. Third result: oddness in DE contexts

Various authors (such as Sauerland 2008 and Percus 2006), have dismissed the hypothesis of an implicature-based theory of oddness based on the observation that it seems to make wrong predictions for the case of embedding in *downward entailing contexts* (DECs). Let me spell out the problem. DECs are contexts, such as sentential negation, ‘no’ or the restrictor of ‘every’, that license inferences such as (21), from the “superset” (‘boy’) to the “subset” (‘tall boy’).

- (21) a. No boy came  $\rightarrow$  No tall boy came.  
 b. Every boy came  $\rightarrow$  Every tall boy came.

It is well known that plain scalar implicatures “flip” in DECs: sentence (17a) with ‘some’ triggers the not-all implicature while sentence (17b) with ‘all’ triggers no implicature; when embedded under negation as in (22), the opposite pattern arises: sentence (22b) with ‘all’ triggers the implicature that John indeed did some of the homework; while sentence (22a) with ‘some’ triggers no implicature.

- (22) a. It is not the case that John did *some* of the homework. = *not(17a)*  
 b. It is not the case that John did *all* of the homework. = *not(17b)*

If oddness comes about through mismatching implicatures, then oddness is expected to flip in DEC: when embedded in a DEC, an odd sentence should turn fine, and vice versa the fine alternative should turn odd. The case of sentential negation shows that this prediction seems not to be born out: sentence (19a), repeated in (23a), remains odd when embedded under negation, as illustrated in (23b).

- (23) a. #Some Italians come from a warm country.  
 b. #It is not the case that some Italians come from a warm country. = *not(23a)*

As my third main result, I have shown in [pub 29] how to properly extend the implicature-based theory of oddness to DEC and how to make sense of apparently problematic examples such as (23). The proposal furthermore offers a new argument that scalar implicatures can be computed in embedded position.

To get started, let's look at the behavior of oddness in DEC different from sentential negation, such as the restrictor of 'every'. Surprisingly, the facts split into two opposite patterns, illustrated by the two cases in (24) and (25). Both pairs (24) and (25) consist of two contextually equivalent scalar alternatives, that only differ on whether the restrictor of 'every' has 'some' or 'all'. I will thus schematically notate the two alternatives as 'EVERY(SOME)' and 'EVERY(ALL)'. In (24), oddness flips as expected: the logically stronger sentence (24b) sounds fine; and the logically weaker sentence (24a) sounds odd. Surprisingly, we get the opposite pattern in (25): it is the logically weaker sentence (25a) that sounds fine; while the logically stronger sentence (25b) sounds odd.

- (24) *Context:* Every year, the dean has to decide: if the college has made enough profit that year, he gives a prize to every professor who has assigned an A to at least some of his students; if there is not enough money, then no one gets a prize.  
 a. #This year, every professor who assigned an A to *all* of his students got a prize. #EVERY(ALL)  
 b. This year, every professor who assigned an A to *some* of his students got a prize.  $\sqrt{\text{EVERY(SOME)}}$
- (25) *Context:* In this department, every professor assigns the same grade to all of his students.  
 a. This year, every professor who assigned an A to *all* of his students got a prize.  $\sqrt{\text{EVERY(ALL)}}$   
 b. #This year, every professor who assigned an A to *some* of his students got a prize. #EVERY(SOME)

What is the relevant difference between the two patterns (24) and (25)? In both cases, the two matrix alternatives 'EVERY(SOME)' and 'EVERY(ALL)' are contextually equivalent, notated ' $=_{ck}$ ' in (26). Yet, contextual equivalence is achieved in a different way in the two cases. In the case of (25), contextual equivalence is achieved at the embedded level: the context says that the restrictors of 'every' in the two sentences coincide, namely that the set of professors who gave an A to *all* students and the set of professors who gave an A to *some* students are the same set; contextual equivalence of the restrictors of course entails matrix contextual equivalence. The case of (24) is different, as contextual equivalence is not established at the embedded level but only at the matrix level: the context leaves open the possibility that some professors gave an A only to some students, so that the restrictors of 'every' in the two sentences are not contextually equivalent.

- (26) Case (24):  $\text{SOME} \neq_{ck} \text{ALL}$ ,  $\text{EVERY(SOME)} =_{ck} \text{EVERY(ALL)}$ ;  
 Case (25):  $\text{SOME} =_{ck} \text{ALL}$ ,  $\text{EVERY(SOME)} =_{ck} \text{EVERY(ALL)}$ ;

I show that my implementation of the theory of scalar implicatures, as sketched in subsection 1.3.3, is fine grained enough to exploit this crucial difference (26) and thus to account for the complex pattern in (24)-(25). The rough idea is as follows. In the case of (25), embedded contextual equivalence forces an embedded mismatching implicature that rules out the sentence 'EVERY(SOME)' with the logically weaker restrictor, even though the matrix sentence would otherwise be logically stronger. In the case of (24), there is no embedded contextual equivalence, and thus no mandatory embedded mismatching implicature; rather, we only get the matrix implicature, which rules out the sentence 'EVERY(ALL)' which is logically weaker at the matrix level. The case of sentential negation (23) considered in the literature hides the actual behavior of oddness in DEC, since negation has the crucial property that matrix contextual

equivalence holds iff embedded contextual equivalence does, so that the two cannot be dissociated and the behavior displayed in (24), where oddness flips as expected, is not visible.

#### 1.3.4. Fourth result: ILPs as a theorem of the theory of oddness

ILPs have figured prominently in the recent literature because they display a huge number of peculiar properties, most of which seem *prima facie* to have nothing to do with the intuitive characterization of ILPs as denoting permanent properties. One case in point is that of restrictions on bare plural subjects (henceforth: BPSs) of ILPs. These restrictions come in various forms. There are *meaning* restrictions on BPSs of ILPs: for instance, BPSs of ILPs lack the existential construal which is instead available for non-ILPs, as illustrated above in (11). And there are *distributional* restrictions on BPSs of ILPs as well: for instance, Diesing (1992) notes that the BPS ‘Feuerwehrmänner’ (‘firemen’) of the ILP ‘intelligent’ cannot sit at the right of the particles ‘ja doch’ (‘indeed’) in German, contrary to the case of the non-ILP ‘verfügbar’ (‘available’), as illustrated in (27).

- (27) a. ... weil ja doch Feuerwehrmänner verfügbar sind.  
           ... since PARTS firemen available are
- b. #... weil ja doch Feuerwehrmänner intelligent sind.  
           ... since PARTS firemen intelligent are

As the fourth main result of my work on ILPs, I have shown that meaning and distributional restrictions on BPSs of ILPs can be derived as a theorem of the general theory of oddness based on mismatching scalar implicatures sketched above. I will not present the details of the account here. But I would like to stress two crucial properties of the proposed account. The *first* property is that my account does not make use of any *ad hoc* grammatical characterization of ILPs. Instead, my account is able to derive these surprising restrictions on BPSs of ILPs from the mere intuitive assumption that ILPs denote (quasi) permanent properties. The *second* property of my account is that it predicts that these restrictions on BPSs of ILPs are waived when the BPS is embedded under a universal operator (over individuals). Fox’s effect (14) above shows that this surprising prediction is borne out for the case of meaning restrictions on BPSs of ILPs. I have shown that this surprising prediction is also borne out for the case of distributional restrictions on BPSs of ILPs. Sentence (28a), with the ILP ‘verwandt’ (‘related to’) and the BPS ‘jüdische Frauen’ (‘Jewish women’) at the right of ‘ja doch’ is deviant, as expected given the contrast in (27). Yet, sentence (28b), where the definite object ‘Chomsky’ has been replaced by the universal quantifier ‘jedem jüdischen Mann’ (‘every Jewish man’), sounds fine.

- (28) a. #Mit Chomsky sind ja doch jüdische Frauen verwandt.  
           with Chomsky are PARTS Jewish women related
- b. Mit jedem Jüdischen Mann sind ja doch jüdische Frauen verwandt.  
           with every Jewish man are PARTS Jewish women related

To the best of my knowledge, my proposal is unique in being able to derive these predictions (14) and (28) concerning embedding of BPSs of ILPs under universal operators.

### 1.4. Main results on error-driven learning and the mode of constraint interaction

*Constraint-based* phonology assumes that speakers are equipped with a universal set of phonological *constraints*, which extract the relevant properties of phonological structures. Two or more constraints can conflict. Languages differ in how they resolve constraint conflicts. Two main modes of constraint interaction have been considered in the literature. According to *Optimality Theory* (OT; Prince and Smolensky 2004), constraints are *ranked* and constraint conflicts are resolved by a language in favor of the constraints that it ranks at the top. According to *Harmonic Grammar* (HG; Legendre, Miyata, and Smolensky 1990b,a; Smolensky and Legendre 2006), constraints are *weighted* and constraint conflicts are resolved by a language additively in terms of the weighted averages of constraint violations.

The choice between OT and HG is an empirical issue which should in the end be decided by typological considerations. Yet, recent research has also started to compare the two frameworks from the perspective of their learnability implications (Riggle 2009; Bane et al. 2010; Jesney and Tessier 2011). Error-driven learning has played a crucial role in this debate. An *error-driven* learner entertains a current OT or HG grammar which is slightly updated whenever it makes an error on the current piece of data, gradually converging to the target grammar. Pater (2008) provides an elegant counterexample against Boersma’s

(1998) GLA implementation of OT error-driven learning. He then diagnoses that “one reason for the failure [of the GLA on his counterexample], and potentially for its failure to provably converge in general, is its use of [...] OT as a model of grammar.” He then notes that “the GLA closely resembles Rosenblatt’s (1958) Perceptron”, which instead uses the HG model of grammar. And he thus concludes advocating the replacement of OT with HG and the adoption of the Perceptron update rule for HG error-driven learning. Pater (2009, p. 1021) indeed argues that “one broad argument for weighted constraints [...] is that weighted constraints are compatible with existing well-understood algorithms for learning variable outcomes and for learning gradually [...]” Additive models of constraint interaction have indeed become widespread in the recent literature (Hayes and Wilson 2008; Boersma and Pater 2007, to appear; Potts et al. 2010; Jesney and Tessier 2011; Coetzee and Kawahara 2013).

In [pub 24], [pub 31], [pub 34], [pub 38], [pub 39], and [pub 40], I have reassessed this debate. I have shown that the problem with Pater’s counterexample is not germane to OT but due to a specific GLA’s implementation detail. Building on previous work (Tesar and Smolensky 1998), I have shown that error-driven learning in OT admits robust computational guarantees which follow from the formal properties of the mode of constraint interaction and thus hold for any constraint set. Finally, I have shown that error-driven learning in HG instead is inefficient without substantive restrictions on the constraint set, thus effectively turning Pater’s assessment on its head. This section sketches the main results I have obtained so far.

#### 1.4.1. First result: a complete computational theory of OT error-driven learning

As recalled above, an *error-driven* learner entertains a current grammar which is slightly updated whenever it makes an error on the current piece of data. The *computational theory* of error-driven learning provides guarantees for four core requirements: *convergence* (is the number of errors made by the learner finite?), *efficiency* (is the number of errors furthermore small?), *stochastic tolerance* (does the number of errors remain small for the stochastic implementation?), and *noise robustness* (does it remain small in the presence of noisy data?). Throughout this section, I illustrate the gist of the analysis of OT error-driven learning by focusing on convergence and efficiency.

Tesar and Smolensky (1998) develop the first provably convergent OT error-driven learner. Its peculiar property is that it only performs constraint demotion of offending constraints, but no constraint promotion of virtuous constraints. Because of this restrictive update schedule, the predicted learning dynamics is too simple to match the attested complexity of child acquisition paths (see subsection 1.5.1 below). These considerations motivate the following research question: is it possible to devise promotion/demotion EDRA’s that provably converge, without restrictive assumptions on the underlying OT typology?

Here is the idea informally. Following Boersma (1997, 1998), assume that the EDRA entertains a numerical representation of the current ranking, by assigning to each constraint a numerical *ranking value* whose relative size reflects the relative ranking of that constraint. Such a numerical representation of the current ranking allows for a numerical formulation of re-ranking rules: constraint demotion consists in decreasing the current ranking value of offending constraints by a small fixed *demotion amount*, say 1 for concreteness; and constraint promotion consists in increasing the current ranking value of virtuous constraints by a small *promotion amount*. As demotion-only has been shown by Tesar and Smolensky (1998) to have a good convergent behavior, the promotion component of the re-ranking rule should not overwhelm the demotion-component, so as not to disrupt too much its good convergent behavior. This requires a proper calibration of the promotion amount. For instance, if there are three constraints that are demoted and two constraints that are promoted, then the promotion amount should be less than  $3/2$ . In fact, two promotions by less than  $3/2$  lead to an overall promotion which is less than the total demotion of 3, so that indeed constraint promotion does not overall overwhelm constraint demotion.

These heuristic considerations suggest that in the general case, the promotion amount should be smaller than the number of constraints demoted divided by the number of constraints promoted, as stated in (29a). Indeed, I show in [pub 31] that convergence holds in this case, with an error-bound which is asymptotically identical to the one obtained by Tesar and Smolensky for the demotion-only case. This guarantee of efficient convergence follows from the OT mode of constraint interaction and does not require any substantive assumption on the constraint set.

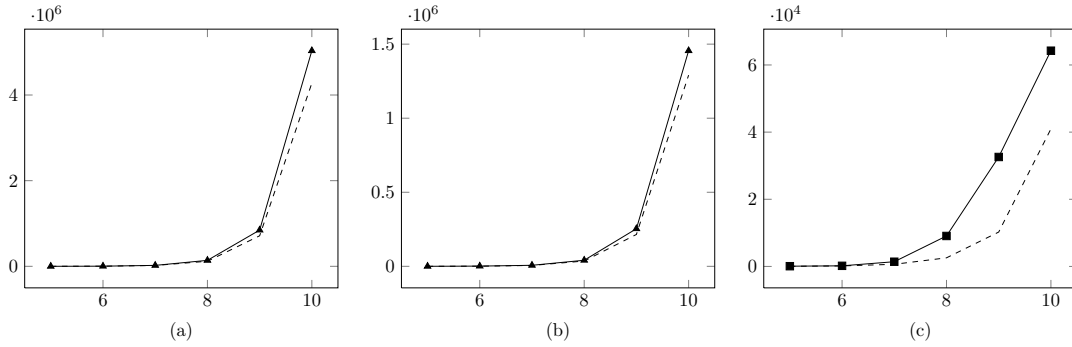


FIGURE 1. Horizontal axis: number  $n$  of constraints (between  $n = 5$  and  $n = 10$ ). Vertical axis: (a) number of errors made by the deterministic HG learner; (b) number of additional errors due to the stochastic component; (c) number of updates needed to recover from a single noisy update.

- (29) a. promotion amount  $< \frac{\text{number of constraints demoted}}{\text{number of constraints promoted}} \implies \text{fast and efficient convergence}$
- b. promotion amount  $= \frac{\text{number of constraints demoted}}{\text{number of constraints promoted}} \implies \text{convergence, but exponentially slow}$
- c. promotion amount  $> \frac{\text{number of constraints demoted}}{\text{number of constraints promoted}} \implies \text{no convergence}$

The ratio between the number of demoted constraints and the number of promoted constraints is called the *calibration threshold*. The calibration condition in (29a) requires the promotion amount to be *strictly* smaller than this calibration threshold. What happens if we increase the promotion amount up to the calibration threshold, as in (29b)? I show that in this case convergence is retained but efficiency is not, as the worst-case number of updates grows exponentially with the number of constraints. The analysis I develop rests on a property of error-driven ranking algorithms that is interesting in its own right: these algorithms explore the typology efficiently, in the sense that they can never entertain again a ranking (vector) that had made a mistake at some earlier time. This property is in turn derived from a connection between the notion of OT-consistency and the geometric property of *conic independence*. Finally, Pater's (2008) counterexample shows that also convergence fails for a promotion amount larger than the calibration threshold, as stated in (29c). The results in (29) thus provide a complete theory of convergence and efficiency for error-driven ranking algorithms. In [pub 38], I have extended these results to the other two crucial issues of the computational theory of error-driven learning: I provide tight guarantees that the OT learner tolerates the stochastic implementation and is robust to noise. Also the latter guarantees follow from the OT mode of constraint interaction and do not require any substantive assumptions on the constraint set. The theory of OT error-driven learning has thus been successfully completed from a constraint-independent perspective.

#### 1.4.2. Second result: comparison with HG error-driven learning

Is it possible to develop analogous computational guarantees for the HG implementation of error-driven learning? In [pub 40], I construct a family of counterexamples which provide a negative answer to this question. To illustrate, the solid line in Figure 1a plots the number of errors made by the HG error-driven learner currently used in the literature on the counterexample corresponding to a number  $n$  of constraints between  $n = 5$  and  $n = 10$ . Comparison with the function  $550 \times 6^{n-5}$  plotted by the dashed line shows that the number of errors grows fast (exponentially) with the number  $n$  of constraints. Indeed, the HG learner makes over five million errors when weighting just  $n = 10$  constraints! For comparison, the OT error-driven learner makes around 40 errors when ranking  $n = 10$  constraints. In conclusion, these simulation results show that it is not possible to develop constraint-independent guarantees of efficient convergence for the HG error-driven learner. Any such guarantees will crucially have to be restricted through specific assumptions on the constraint set.



The implementation of error-driven learning considered so far is called *deterministic*, to distinguish it from the *stochastic* implementation (Boersma 1997, 1998; Boersma and Hayes 2001; Coetzee and Pater 2008, 2011; Coetzee and Kawahara 2013; Jarosz 2013; Boersma and Pater to appear). The latter differs because the current piece of data is compared not with the current grammar but with a variant thereof sampled from a neighborhood of the current grammar. In the worst-case scenario, the stochastic learner can make more errors than the deterministic learner, because the stochastic component can derail the learner away from the most straightforward learning path. The solid line in Figure 1b plots the difference between the largest number of errors made by the stochastic HG learner minus the number of errors made by the deterministic HG learner. In other words, it plots the additional number of errors due to the stochastic component. Comparison with the function  $166 \times 6^{n-5}$  plotted by the dashed line shows that the number of additional errors grows fast (exponentially) with the number  $n$  of constraints. Indeed, in the case with just  $n = 10$  constraints, the stochastic learner performs almost one million and a half additional errors compared with the deterministic learner, which in turn was already performing over five million errors! For comparison, the OT error-driven makes less than 200 additional errors when ranking  $n = 10$  constraints. In conclusion, these simulation results show that it is not possible to develop constraint-independent guarantees that the HG error-driven learner tolerates the stochastic implementation, namely that the number of additional errors due to the stochastic component remains small. Any such guarantees will crucially have to be restricted through specific assumptions on the constraint set.

Finally, the solid line in Figure 1c plots the number of updates required by the HG learner to recover from the update triggered by a single corrupted, noisy piece of data (which changes at most three weights by at most 3). Comparison with the function  $40 \times 4^{n-5}$  plotted by the dashed line shows that the number of updates needed to recover grows fast (exponentially) with the number  $n$  of constraints. So fast that the learner requires over 64,000 updates in the case with just  $n = 10$  constraints. In conclusion, this counterexample shows that constraint-independent guarantees of noise-robustness for the HG error-driven learner are impossible, as a single faulty update can require a large number of updates to recover. Any such guarantees will have to be restricted through specific assumptions on the constraints.

### 1.4.3. Third result: algorithmic portability from HG into OT

The peculiar property of OT is that it uses *constraint ranking* and thus enforces *strict domination*, according to which the highest ranked relevant constraint “takes it all”. Because of this property, OT seems *prima facie* not to have any close correspondent within machine learning. Thus, computational OT has developed so far combinatorial algorithms, tailored to the specific OT framework, with little connections to methods and results from computational linguistics or machine learning. As recalled at the beginning of this section, additive models such as HG have thus been singled out in the literature as a way to bridge this gap between computational constraint-based phonology and machine learning. Let me recall again Pater’s (2009; p. 1021) statement: “One broad argument for [“the replacement of OT ranking with HG weighting”] is that weighed constraints are compatible with existing well-understood algorithms for learning variable outcomes and for learning gradually [...]. As these algorithms are broadly applied with connectionist and statistical models of cognition, this forms an important connection between HG [...] and other research in cognitive science.”

In [pub 34], I prove that this conjecture of an alleged computational superiority of HG over OT is false. In fact, I develop a general strategy that allows any algorithm for HG to be ported into OT. It follows that HG has no computational advantage over OT. This result is important because it opens the way to an approach to computational OT radically different from the one pursued so far in the literature: rather than devising from scratch *ad hoc* combinatorial algorithms, computational problems that arise in computational OT can be tackled by importing and straightforwardly adapting well known algorithms from Machine Learning.

So far, I have explored two applications of this new approach to Computational OT. Tesar (1995) develops a batch OT ranking algorithm, called *Recursive Constraint Demotion* (RCD). His analysis of RCD is combinatorial in nature and specifically tailored to the logic of OT. As a *first* illustration of this new approach to computational OT, I show that RCD can be reinterpreted as the classical *Fourier-Motzkin Elimination Algorithm* for solving systems of linear inequalities (Bertsimas and Tsitsiklis 1997). This reinterpretation might turn out useful in improving later applications of RCD such as *Biased RCD* (Prince and Tesar 2004) and *Low Faithfulness RCD* (Hayes 2004). As a *second* illustration of my new approach to computational OT, I derive an alternative very different proof of finite time convergence for the calibrated OT error-driven learner, that hinges on the convergence theorem for the celebrated

*Perceptron Algorithm* for linear classification. This alternative proof is interesting because it can be generalized from the Perceptron to any other online algorithm for linear classification, thus leading to a potentially large number of new OT online algorithms, whose modeling predictions are still completely unexplored.

### 1.5. Main results on restrictiveness and the early acquisition of phonotactics

Knowledge of the *phonotactics* of the target language is knowledge of the distinction between licit and illicit sounds and sound concatenations. For instance, English speakers know that [blik] would be a licit English word while \*[bnik] would not, although both are unattested in the English lexicon, and despite the fact that both would be licit in Arabic. Based on a comprehensive review of the psycholinguistic literature, Hayes (2004) concludes that the acquisition of phonotactics goes through an *early stage* (around one year of life) that can be described by the two (somewhat idealized) properties in (30).

- (30) a. *Properties of the input.* Throughout the early stage, morphology lags behind and the child is thus blind to alternations.
- b. *Properties of the output.* By the end of the early stage, the child is able to distinguish legal from illegal structures, namely displays knowledge of the target adult phonotactics.

Of course, (30) is the statement of a *computational problem*, namely a mapping from an input (30a) to an output (30b). Most of my work in computational phonology (and in particular [pub 16], [pub 18], [pub 19], [pub 20], [pub 23], [pub 32], [pub 33], and [pub 36]) has focused on this problem.

My core research hypothesis is that OT error-driven learning (introduced in the preceding chapter 1.4) provides a computationally sound and cognitively plausible model of the early acquisition of phonotactics. I have so far tried to establish this research hypothesis from two perspectives. One perspective focuses on learnability: under which assumptions on the constraint set and on the stream of learning data, can I guarantee that the final OT grammar entertained by the OT error-driven learner at converge captures the target phonotactics? Another perspective focuses instead on fine grained data on the time-course of acquisition: how well does the OT error-driven learner predict the distribution of child repair strategies? This section sketches the main results I have obtained so far.

#### 1.5.1. First result: constraint promotion is needed from a modeling perspective

As recalled in chapter 1.4, an OT error-driven learner slightly re-ranks the constraints whenever the current ranking makes an error on the current piece of data. Re-ranking can involve two types of operations: offending constraints can be *demoted* to a lower position in the ranking; and virtuous constraints can be *promoted* to a higher position. As recalled, Tesar and Smolensky (1998) develop the first error-driven ranking algorithm, called *Error Driven Constraint Demotion* (EDCD). Its signature property is that it demotes offending constraints but does not promote virtuous constraints. Lack of constraint promotion allows Tesar and Smolensky (1998) to prove that EDCD converges efficiently. Yet, although a virtue from a *computational* perspective, lack of constraint promotion turns out to be a liability from a *modeling* perspective. Informal arguments in favor of constraint promotion have been provided by Gnanadesikan (2004) and Bernhardt and Stemberger (1998); Boersma and Hayes (2001) provide a computational argument, but framed within a stochastic variant of the standard OT framework. In [pub 31], I develop a new formal argument in favor of constraint promotion within standard OT, based on the modeling implications of the early stage of the acquisition of phonotactics.

The gist of the argument can be informally sketched as follows. As summarized in (30a), morphology lags behind throughout this early stage and the learner thus lacks the information on underlying forms provided by morphological alternations. The best the learner can do is thus to assume fully faithful underlying forms (Prince and Tesar 2004; Hayes 2004). This means that the faithfulness constraints never make any mistake and are thus never re-ranked by a demotion-only ranking algorithm such as EDCD. This cannot be right. In fact, if two phonotactic patterns in the typology require the opposite relative ranking of some faithfulness constraints, EDCD fails on at least one of them (see below for examples of such cases). Furthermore, EDCD is unable to model learning paths where the child's repair strategy for a certain marked structure changes over time (for instance McLeod et al. 2001 document leaning paths where complex onsets are simplified first by deletion and then by coalescence). In conclusion, EDCD does not provide an adequate implementation of error-driven learning and some amount of constraint promotion is needed.

### 1.5.2. Second result: intractability of the problem of the acquisition of phonotactics

The research summarized so far focuses on the *input property* (30a) of the early stage: I have formalized lack of alternations through the assumption that the learner posits faithful underlying forms; and I have explored the consequences of this assumption for the choice of proper re-ranking rules. Let me now turn to the *output property* (30b) of the early stage: despite lack of alternations, the learner does acquire (most of) the target phonotactics. Knowledge of phonotactics is twofold. On the one hand, it requires knowing which structures are *consistent* with the target phonotactics, such as [blik] in English. On the other hand, it requires knowing which structures are *inconsistent* with the target phonotactics, such as \*[bnik]. These two components are formalized in the literature through the *consistency* and *restrictiveness* conditions in (31a) and (31b), respectively (Berwick 1985; Manzini and Wexler 1987; Prince and Tesar 2004; Hayes 2004).

- (31) *Given:* a finite set of surface forms all phonotactically licit according to the phonotactics of some grammar in the underlying typology.  
*Find:* a grammar in the underlying typology which is:
- a. *consistent*, namely it predicts each training surface form to be phonotactically licit;
  - b. *restrictive*, namely there exists no other grammar in the typology which is consistent with the training data as well and furthermore predicts a subset of forms to be phonotactically licit.

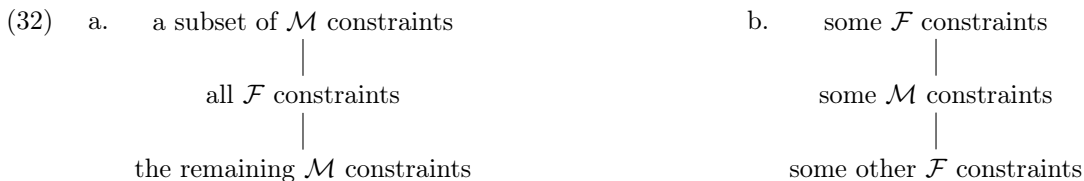
This problem (31) formalizes the challenge raised by the early acquisition of phonotactics.

Convergence requires the OT error-driven learner to make only a finite number of errors. Any convergent learner thus returns a *final grammar* which satisfies the consistency condition (31a) — otherwise, the learner would still make an error on the training data, contradicting the hypothesis that it has converged. As recalled in subsection 1.4.1, efficient convergence can be established for OT error-driven learners without any restrictive assumptions on the underlying constraint set. In other words, the consistency condition (31a) by itself is easy to meet. Unfortunately, the addition of the restrictiveness condition (31b) drastically changes the quality of the problem. In fact, I prove in [pub 7] and [pub 33] that the problem (31) of the acquisition of phonotactics within OT is *intractable*: it cannot be solved efficiently by any algorithm, no matter whether it is error-driven or not. In the sense that for any alleged solution algorithm, there exists an instance of the problem for which that algorithm fails. The proof is by polynomial-time reduction from the Cyclic Ordering problem, a well-known NP-complete computational problem.

“Discovering that a problem is [intractable] is to arrive at the beginning of a computational challenge, rather than at the end of an inquiry” (Clark and Lappin 2011, p. 139). Indeed, Barton et al. (1987, p. 4) explain that “[intractable] problems don’t have any special structure that would support an efficient solution algorithm, so there’s little choice but brute force”. These problems “might well be characterized as *unnatural*” because “there is every reason to believe that natural language has an intricate computational structure that is not reflected in combinatorial search methods”. They thus conclude that an intractable problem “leaves unmentioned some constraints of the natural problem.” From this perspective, the complexity result just mentioned motivates the following question: which restrictions provide sufficient additional structure to develop efficient solution algorithms for the problem (31)? Restrictions come in two types: restrictions on the target phonotactic patterns and restrictions on the constraint set, leading to the two sets of results described in subsections 1.5.3 and 1.5.4.

### 1.5.3. Third result: restrictiveness through assumptions on phonotactic patterns

As the very first step in the acquisition of phonotactics, the child needs to learn the inventory of licit *segments* (the basic phonological units) in the target adult language. When modeling a segment inventory in OT, it often happens that some subset of the markedness constraints punish exactly all and only the illicit segments. In this case, the inventory can be modeled through a ranking such as (32a). The designated subset of markedness constraints hold sway at the top of the ranking while the remaining markedness constraints are silent at the bottom. The relative ranking of the faithfulness constraints sandwiched in between is irrelevant. The inventory can thus be called  *$\mathcal{F}$ -irrelevant*, to distinguish it from inventories which instead require some faithfulness constraints to be crucially ranked above some other faithfulness constraints in order to sandwich some markedness constraints in between, as illustrated in (32b).



For instance, suppose that the constraint set contains a markedness constraint against voiced velar obstruents and another markedness constraint against dorsal fricatives. The velar inventory [g k ɣ x] which only consists of the voiceless velar stop (illicit segments are stricken out) simply requires those two markedness constraints to hold sway at the top, as they together punish the three illicit segments [g], [ɣ], and [x]. These considerations trivially extend from segment inventories to arbitrary phonotactic patterns.

In [pub 16], [pub 19], [in prep 1], [in prep 2], and [in prep 3], I have obtained the first result on correctness of OT error-driven ranking algorithms. Roughly stated, this result says that, whenever the learner does not perform “too much” constraint promotion, it succeeds at learning any target phonotactic pattern which is  $\mathcal{F}$ -irrelevant, namely which does not require any specific relative ranking of the faithfulness constraints. This result holds under no assumptions on the constraint set, apart from the mild assumption that the generating function be *symmetric*, in the sense that  $y$  is a candidate for  $x$  if and only if  $x$  is vice versa a candidate for  $y$ . This learnability result does not extend to OT error-driven learners which promote “too much”, such as Boersma’s (1998) *Gradual Learning Algorithm* (GLA). This shows that success on  $\mathcal{F}$ -irrelevant languages is not trivial, and requires a proper choice of the implementation details. As tested on a number of specific test cases,  $\mathcal{F}$ -irrelevant languages make up the large majority of OT typologies. The intuitive reason is that, as the relative ranking of the faithfulness constraints mainly governs the repair strategies, it turns out to be crucial for alternations, but mostly irrelevant for phonotactics. As  $\mathcal{F}$ -irrelevant languages represent the vast majority of any typology, correctness of OT error-driven learning on these languages represents a substantial result.

#### 1.5.4. Fourth result: restrictiveness through assumptions on constraint sets

The result just recalled copes with the intractability of the problem (31) of the acquisition of OT phonotactics through a restriction on the target phonotactic patterns (namely, the  $\mathcal{F}$ -irrelevance assumption). Another approach is to restrict the constraint sets which define the underlying OT typology. Focusing again on the acquisition of segment inventories, I assume that segments are described through total, binary phonological features (roundness, voicing, height, etcetera) that extract the relevant segmental properties. Each feature comes with a dedicated faithfulness constraint and a dedicated markedness constraint. Besides these unary constraints, there is a set of *feature co-occurrence constraints* (FCCs), which penalize certain combinations of feature values. The learning algorithm needs to infer a ranking of these constraints that corresponds to the inventory of licit segments in the target adult phonotactics. The learner is required to succeed no matter whether the target inventory is  $\mathcal{F}$ -irrelevant or not.

In [pub 10] and [pub 23], I start to investigate the conditions under which an OT error-driven learner is able to tackle this task, obtaining the following preliminary result. If the degree of feature interaction is limited (i.e., each feature interacts with at most another feature) and the mode of feature interaction is phonologically plausible (i.e., no FCC punishes a form which has the unmarked value for each feature targeted by that FCC), then the OT error-driven learner with a calibrated promotion component (as defined in subsection 1.4.1) succeeds on any phonotactic pattern in the typology (no matter whether it is  $\mathcal{F}$ -irrelevant or not), no matter what the input stream of data looks like. Interestingly, this learnability result does not extend to two other OT learners available in the current literature, namely Tesar and Smolensky’s (1998) EDCD and Boersma’s (1997) GLA.

I am currently extending these preliminary results along the following lines. Towards the goal of developing a comprehensive formal theory of FCCs, I want to explore the *Tree Hypothesis* (TH). It maintains that the FCCs define universal entailments among the features which are representable through a *feature interaction graph* which crucially has no loops, namely it is a tree. I want to explore the TH from the perspectives of learnability and typology. My first goal is to show that the TH provides enough structure to derive restrictiveness guarantees for a properly designed OT error-driven learner—while OT phonotactics is provably unlearnable without a substantive markedness theory, as recalled above. My second goal is to develop a system of FCCs which complies with the TH and has good typological coverage relative to various available databases of segment inventories.

### 1.5.5. Fifth result: modeling child epenthesis data from the INANP database

The acquisition of phonotactics is characterized by a gradual, stepwise progression towards the target adult grammar. The OT acquisition literature has endorsed the error-driven learning model because it has the potential to model these child acquisition paths, as it describes a sequence of current rankings, each corresponding to an intermediate phonotactics. The research summarized so far has contributed algorithmically sound implementations of the error-driven model. Yet, this computational perspective needs to be complemented by an investigation of the modeling implications of these algorithms, focusing on the following core research question: is it the case that the learning sequences formally predicted by these computationally sound implementations of the error-driven ranking model match attested child acquisition paths?

In collaboration with Adam Albright (MIT), we have focused on the database collected as part of the *Iowa-Nebraska Articulation Norms Project* (INANP; Smit et al. 1990), that I have acquired from Ann Bosma Smit at Kansas State University. It consists of transcribed productions of all English singleton consonants in both onset and coda position, as well as of the 27 most common English onset (bi- and tri-) consonant clusters, for a total of 108 targets per child. The database contains production data from more than 500 English-learning, normally developing children with age between 2 and 9 years. Child productions were collected and coded by trained speech pathologists. The database had so far only been used within the speech-pathology literature, to determine articulation norms for the diagnosis of speech impairment. The database had not been the subject of phonological analysis and its modeling implications were still completely unexplored. We have used this database for the first time for phonological and computational modeling. The dimension of the database by far exceeds the datasets that have been considered so far in the OT child phonology and modeling literature, for the number of children (over 500), the wide range of ages considered (from 2 to 9 years) and the exhaustivity of the set of phonological targets (virtually all English onset consonant clusters).

We have focused on child’s patterns of epenthesis into consonant clusters ( $/CCV/ \rightarrow [CVCV]$ ). In principle, this process may be motivated by a strong articulatory preference in children for mandibular oscillation, favoring CV sequences (MacNeilage 1998); or it may be motivated by the same phonological constraints that derive epenthesis in adult phonologies. We provide evidence that epenthesis in child English is not merely a result of articulatory pressures, but is shaped by the same set of perceptually motivated constraints that govern epenthesis in adult phonologies. Here is the gist of the argument. In adult systems, converging evidence from reduplication, infixation, loanword adaptation, alliteration, and puns shows that epenthesis is preferred in stop+liquid clusters ( $/pra/ \rightarrow [pVra]$ ), relative to s+stop clusters ( $/sta/ \rightarrow [sVta]$ ). Fleischhacker (2001, 2005) attributes this asymmetry to the greater perceptual similarity of  $[pra] \sim [pVra]$ , and the lesser similarity of  $[sta] \sim [sVta]$ . Based on data from the INANP database, we show in **[in prep 4]** that children are subject to the very same set of asymmetries. This finding supports the strong *continuity hypothesis* that children possess the same set of representations and constraints as adults.

This finding furthermore yields a new approach to the long standing problem posed by children who acquire s+stop clusters before other cluster types (Barlow 2001; Fikkert 1994), despite the fact that s+stop clusters are marked and are thus expected to be acquired later. Approaches that posit a special status for the initial  $/s/$  are unable to account for children that acquire s+stop clusters before other sC clusters. Our findings pave the way for a new approach. As just recalled, epenthesis into s+stop clusters is heavily dispreferred in adult phonology. Also deletion has been reported to be dispreferred in the case of s+stop clusters (Fleischhacker 2005). Our findings show that (at least some of) these dis-preferences for certain repair strategies for certain cluster types carry over to child phonology. This suggests the following approach to the precocious acquisition of s+stop clusters in certain developmental paths: s+stop clusters are acquired early despite their marked status because they are “harder to simplify”, i.e. epenthesis and deletion incur a higher cost (say, a violation of a higher ranked faithfulness constraint). We explicitly implement this idea within the error-driven learning scheme using Fleischhacker’s family of DEP-constraints.

## Project overview

According to *generative phonology*, phonological grammars map *underlying forms* to corresponding *surface forms* (Kenstowicz and Kisseberth 1977; Heinz 2011). This work looks at two structural properties of phonological grammars: *idempotency*, which requires phonotactically licit forms to be faithfully mapped to themselves (Hayes 2004; Prince and Tesar 2004); and *output-drivenness*, which requires any discrepancy between underlying and surface (or *output*) forms to be driven exclusively by the phonotactics (Tesar 2013). Output-drivenness entails idempotency: if unfaithful mappings are only motivated by the phonotactics (as demanded by output-drivenness), then phonotactically licit forms cannot surface unfaithfully (as required by idempotency).

This work extends and systematizes the theory of idempotency and output-drivenness within *constraint-based* phonology (Prince and Smolensky 2004). This framework assumes that a set of *constraints* extract the relevant properties of phonological representations, analogously to *features* in machine learning (Schölkopf and Smola 2002) and cognitive science (since Shepard et al. 1961). Constraints come in two types: *markedness* constraints, which measure the phonotactic badness of surface forms; and *faithfulness* constraints, which measure the disparity between underlying targets and corresponding surface realizations. Two implementations of constraint-based phonology are considered, which differ for the mode of constraint interaction. *Optimality Theory* (OT; Prince and Smolensky 2004) assumes constraints to be ranked and thus resolves constraint conflicts in favor of the single top ranked relevant constraint. *Harmonic Grammar* (HG; Legendre et al. 1990b,a; Smolensky and Legendre 2006) instead assumes constraints to be weighted and thus resolves constraint conflicts in terms of the weighted average of constraint violations.

Within constraint-based phonology, the crucial question of the theory of idempotency and output-drivenness can be stated as follows: which conditions on the constraint set ensure that all the grammars in the corresponding OT or HG typology satisfy the structural properties of idempotency or output-drivenness? Two main findings emerge in addressing this question. First, all available sufficient conditions for idempotency and output-drivenness only target the faithfulness constraints: surprisingly, they place no restrictions whatsoever on the markedness constraints. Second, the *basic* faithfulness constraints posited in the phonological literature all satisfy these sufficient conditions for idempotency and output-drivenness: surprisingly, the only faithfulness constraints which fail at these sufficient conditions are those that are *derived* from the basic ones through either constraint *aggregation* (e.g., constraint conjunction) or constraint *restriction* (e.g., restriction to certain specific positions, to certain specific segments, etcetera). These findings are rather surprising: given that the faithfulness constraint conditions for idempotency and output-drivenness all look technical, rather bizarre and hard to interpret, why is it the case that the basic faithfulness constraints posited in the phonological literature all happen to satisfy them?

The main contribution of this work is a solution to this puzzle: despite first appearance, the sufficient conditions for idempotency and output-drivenness are shown to all admit a uniform, intuitive interpretation which explains why the basic faithfulness constraints all satisfy these conditions. Here is the core intuition. Faithfulness constraints measure the phonological distance between underlying and surface forms. They should thus comply with a crucial axiom of the definition of distance (or *metric*), namely that any side of a triangle is shorter than the sum of the other two sides. This intuition can be straightforwardly formalized through a *faithfulness triangle inequality*. This inequality is in turn shown to be equivalent to the various faithfulness sufficient conditions for idempotency and output-drivenness. These equivalences hold under various additional assumptions, crucially including McCarthy’s (2003b) generalization that faithfulness constraints are all *categorical*. In conclusion, the reason why the basic faithfulness constraints posited in the phonological literature all satisfy the sufficient conditions for idempotency and output-drivenness is that these conditions are equivalent to the metric triangle inequality and the basic faithfulness constraints all comply with this inequality because they measure phonological

distance in a sensible way. This initial chapter provides an informal overview of the results obtained, roughly section by section.

The connection between idempotency/output-drivenness and the metric properties of the faithfulness constraints has implications for phonological theory. A grammar fails at idempotency provided it displays a *chain shift* (Łubowicz 2011, Moreton and Smolensky 2002, and reference therein). An idempotent grammar fails at output-drivenness provided it displays a *derived environment effect* or a *saltation* (Łubowicz 2002, White 2013 and references therein). Thus, output-drivenness captures the classical notion of *transparency* developed in the theory of rule ordering: failure at output-drivenness corresponds to counter-feeding opacity (chain shifts) or counter-bleeding opacity (saltations). It is often claimed in the literature that constraint-based phonology is incompatible with opacity because of its non-derivational architecture and its related lack of intermediate representational levels. As a recent example, Kawahara (2006) writes: “phonological opacity involves a generalization that cannot be stated solely by reference to surface structures. The classic, non-derivational version of OT does not predict the existence of phonological opacity, as it is surface-oriented.” The theory presented here shows that this alleged “surface-orientedness” (formally captured by output-drivenness) does not follow from lack of derivations and intermediate representational levels. Indeed, it is not an architectural property of constraint-based phonology. Instead, surface-orientedness/output-drivenness requires a specific restriction on the theory of faithfulness: that all faithfulness constraints measure phonological distance in a sensible way, namely in compliance with the metric triangle inequality.

Finally, the theory of opacity presented in this work has implications for learnability. The core idea of computational generative linguistics is that “for language learning to be possible, the linguistic theory must have some kind of non-trivial structure [...] that can be exploited by a learner” (Tesar 2013, p. 18). The structure provided by idempotency and output-drivenness has figured prominently in the literature on the learnability of phonology. The typological structure provided by idempotency has been argued to boost the learnability of phonotactics, as it allows the learner to safely assume a faithful underlying form for each phonotactically licit training surface form (Gnanadesikan 2004; Hayes 2004; Prince and Tesar 2004). And Tesar (2013) argues that the typological structure provided by output-drivenness boosts the learnability of underlying forms for a given set of training surface forms. Sufficient conditions for idempotency and output-drivenness thus provide solid computational guarantees for various current learning models.

### 2.1. Idempotency and counter-feeding opacity (section 3.1)

Suppose there are no representational differences between underlying and surface forms: any given representation can be construed as both an underlying and a surface form (see Moreton 2004b for discussion). In this case, a phonological grammar is called *idempotent* (Hayes 2004; Prince and Tesar 2004) provided any form which is phonotactically licit (as a surface form) is faithfully mapped (as an underlying form) to itself (as a surface form). Section 3.1 formalizes this notion of idempotency within a representational framework where underlying and surface forms are strings of segments related by *correspondence relations* in the sense of McCarthy and Prince (1995).

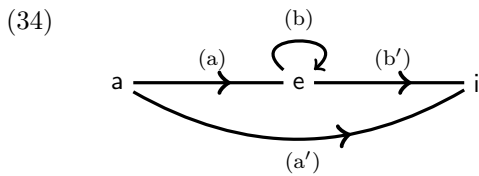
To appreciate the phonological relevance of the notion of idempotency, consider a phonological grammar which fails at idempotency. This means that there exists some phonological form (say the mid vowel *e*) which is phonotactically licit according to that grammar and yet it is not faithfully realized. The assumption that [e] is phonotactically licit means that it is “attainable” by the grammar, in the sense that it is the surface realization of some underlying vowel, say of the low vowel /a/ for concreteness. The assumption that the mid vowel /e/ is not faithfully mapped to itself means that it is realized as some other vowel, say as the high vowel [i] for concreteness. In conclusion, the grammar considered fails at idempotency because it enforces the *chain shift* (34).

$$(33) \quad a \longrightarrow e \longrightarrow i \overset{\curvearrowright}{\longleftarrow}$$

Within SPE, chain shifts are the result of *counter-feeding* rule interactions (Baković 2011 and references therein). In conclusion, the notion of idempotency is relevant to phonology because its negation captures counter-feeding opacity.

## 2.2. The faithfulness idempotency condition (section 3.2)

Within constraint-based phonology, structural properties of a typology of grammars follow from properties of the constraint set.<sup>1</sup> I focus on a *classical* implementation of *Optimality Theory* (OT; Prince and Smolensky 2004), whereby the constraint set only consists of markedness and faithfulness constraints. Which assumptions on the constraint set suffice to guarantee that the OT grammar corresponding to any constraint ranking is idempotent? Here is a preview of the answer developed in section 3.2 (this reasoning was anticipated in Moreton and Smolensky 2002, section 3, Prince 2007, and Buccola 2013; furthermore, it is analogous to the analysis in Tesar 2013, section 3.2). Suppose that an OT grammar maps the underlying form /a/ to the surface form [e], as represented by the arrow (34a). This means that [e] is phonotactically licit. Idempotency then requires the underlying form /e/ to be faithfully mapped to [e], as represented by the loop (34b).



We reason by contradiction. Thus, we make the *contradictory assumption* that idempotency fails and that /e/ is instead mapped to something else, say it is raised further to [i] for concreteness, as represented by the arrow (34b'). In order to establish idempotency, we want to derive the *contradictory conclusion* that /a/ is also mapped to [i], as represented by the long arrow (34a'), against the hypothesis that /a/ be mapped to [e].

Assume that every constraint  $C$  in the constraint set satisfies the implication (35). The contradictory assumption that /e/ is raised to [i] rather than faithfully mapped to [e] intuitively means that high ranked constraints fail at penalizing the contradictory mapping (/e/, [i]) in (34b') with respect to the idempotent mapping (/e/, [e]) in (34b), thus satisfying the antecedent of (35).

- (35) **If:** constraint  $C$  does not prefer the idempotent mapping (/e/, [e]) to the contradictory mapping (/e/, [i]) *namely:*  
 $C(/e/, [i]) \leq C(/e/, [e])$
- Then:** constraint  $C$  does not prefer the actual mapping (/a/, [e]) to the contradictory mapping (/a/, [i]) *namely:*  $C(/a/, [i]) \leq C(/a/, [e])$

The implication (35) thus ensures that high ranked constraints also fail at penalizing the contradictory mapping (/a/, [i]) in (34a') with respect to the actual mapping (/a/, [e]) in (34a). In conclusion, we intuitively expect this implication (35) to provide a sufficient condition for the contradictory assumption to entail the contradictory conclusion, thus guaranteeing the idempotent mapping of /e/ to [e].

The mappings (/e/, [e]) and (/e/, [i]) compared in the antecedent of (35) feature the underlying form /e/. The mappings (/a/, [e]) and (/a/, [i]) compared in the consequent only differ because they feature the underlying form /a/. The implication (35) thus trivially holds for the markedness constraints, because they are insensitive to the underlying forms, so that antecedent and consequent coincide. The implication (35) is thus a condition on the faithfulness constraints. For a faithfulness constraint, the number of violations assigned to the identity mapping (/e/, [e]) is zero. The implication (35) therefore becomes (36), where I have replaced “ $C$ ” with “ $F$ ”, to highlight the fact that the implication only needs to be checked for the faithfulness constraints.

- (36) **If:**  $F(/e/, [i]) = 0$
- Then:**  $F(/a/, [i]) \leq F(/a/, [e])$

Section 3.2 formalizes this reasoning and concludes that idempotency holds for the OT grammar corresponding to any ranking of a given constraint set provided every faithfulness constraint  $F$  in the constraint set satisfies condition (37) for any forms  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$ . This condition is thus referred to as the *faithfulness idempotency condition* (FIC).

<sup>1</sup> Properties of the candidate set also play a crucial role in shaping a typology in constraint-based phonology. This introductory chapter offers only an informal preview and thus omits various candidate conditions which will be carefully discussed in the following chapters.



$$(37) \quad \boxed{\begin{array}{l} \mathbf{If:} \quad F(\mathbf{b}, \mathbf{c}) = 0 \\ \mathbf{Then:} \quad F(\mathbf{a}, \mathbf{c}) \leq F(\mathbf{a}, \mathbf{b}) \end{array}}$$

Here is another way to appreciate the intuitive meaning of the FIC. Recall that idempotency fails in the presence of a chain shift such as  $\mathbf{a} \rightarrow \mathbf{e} \rightarrow \mathbf{i}$ . In order to get the chain shift  $\mathbf{a} \rightarrow \mathbf{e} \rightarrow \mathbf{i}$ , the top ranked relevant faithfulness constraint cannot penalize the mapping  $(/e/, [i])$  relative to  $(/e/, [e])$ , thus satisfying the antecedent of the FIC (36). Furthermore, it cannot penalize the mapping  $(/a/, [e])$  relative to  $(/a/, [i])$ , thus failing at the consequent of the FIC (36).

### 2.3. Which faithfulness constraints satisfy the FIC? (sections 3.3-3.5)

Given the sufficient condition for idempotency provided by the FIC (37), the problem of establishing OT idempotency is thus reduced to the problem of determining which faithfulness constraints satisfy the FIC. The latter problem is taken on by sections 3.3-3.5, for a variety of faithfulness constraints which naturally arise within McCarthy and Prince’s (1995) *correspondence theory* of faithfulness and its recent developments. To start, section 3.3 looks at the three basic faithfulness constraints MAX, DEP, and IDENT. MAX is shown to satisfy the FIC under no additional assumptions while DEP and IDENT require no correspondence relation in the candidate set to break any underlying segment. This edge of MAX over DEP and IDENT can be intuitively explained as follows. The left- and right-hand side in the consequent of the FIC (37) involve two mappings which share the underlying string (namely  $\mathbf{a}$ ) but differ for the surface strings (namely  $\mathbf{b}$  and  $\mathbf{c}$ ). While MAX only “counts” over underlying segments, DEP and IDENT are also sensitive to the surface segments and thus need some additional assumption (in the form of the no-breaking condition) to guarantee the “commensurability” of the two different surface forms in the two candidates being compared.

Section 3.4 extends the analysis of the FIC to *restricted* variants of these basic constraints, such as faithfulness constraints which punish consonant deletion but not vowel deletion, which punish denasalization but not nasalization, or which punish obstruent devoicing only before a sonorant. These restricted faithfulness constraints are shown to fail at the FIC as soon as the correspondence relations are allowed to *cross* the restriction, namely to establish a correspondence between a segment which does satisfy the restriction and one which does not. For instance, a variant of MAX restricted to consonants (it does not penalize vowel deletion) fails at the FIC whenever the candidate set contains candidates where a consonant is in correspondence with a vowel.

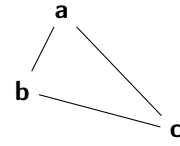
Section 3.5 concludes the analysis of the FIC by looking at a variety of other faithfulness constraints which naturally arise within correspondence theory, such as INTEGRITY, UNIFORMITY, featural DEP $_{[\pm\varphi]}$  and MAX $_{[\pm\varphi]}$ , CONTIGUITY, ALIGNMENT, and LINEARITY. Constraints obtained by aggregation (such as conjunction and disjunction) of basic constraints are investigated as well. The overall picture that arises from this investigation is that all *basic* faithfulness constraints do satisfy the FIC (at least when correspondence relations are not allowed to break underlying segments). The constraints which fail at the FIC are those *derived* from basic ones, through either constraint restriction or constraint aggregation.

### 2.4. The faithfulness triangle inequality (section 4.1)

Idempotency has been made to follow from the sufficient condition provided by the FIC in (37). This condition looks admittedly rather technical, without a straightforward interpretation. One might thus expect only a small subset of the faithfulness constraints to satisfy this apparently bizarre condition. Yet, sections 3.3-3.5 have shown that all *basic* faithfulness constraints used in the OT literature seem to satisfy it. Indeed, the only faithfulness constraints which fail at it are those *derived* from basic ones, either through constraint aggregation (such as constraint conjunction) or restriction. The fact that the basic faithfulness constraints all happen to satisfy the FIC is puzzling and calls for an explanation. Obviously, it must be the case that the FIC captures some deep formal condition that phonologists have implicitly assumed to be crucial in designing proper faithfulness constraints. What is this condition?

Intuitively, faithfulness constraints measure the “distance” between underlying and surface forms along various phonologically relevant dimensions. It thus makes sense to investigate whether faithfulness constraints satisfy formal properties of distances (or *metrics*). One such important property is the *triangle inequality* (105): it captures the intuition that the distance between any two points  $\mathbf{a}$  and  $\mathbf{c}$  is shorter than the distance between  $\mathbf{a}$  and  $\mathbf{b}$  plus the distance between  $\mathbf{b}$  and  $\mathbf{c}$ , for any choice of the intermediate point  $\mathbf{b}$  (Rudin 1953).

$$(38) \quad \text{distance}(\mathbf{a}, \mathbf{c}) \leq \text{distance}(\mathbf{a}, \mathbf{b}) + \text{distance}(\mathbf{b}, \mathbf{c})$$



Based on these considerations, section 4.1 says that a faithfulness constraint  $F$  satisfies the *faithfulness triangle inequality* provided condition (107) holds for three phonological forms  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  (recall that I am assuming that there are no representational differences between underlying and surface forms, so that any form can figure as both an underlying and a surface form).

$$(39) \quad F(\mathbf{a}, \mathbf{c}) \leq F(\mathbf{a}, \mathbf{b}) + F(\mathbf{b}, \mathbf{c})$$

The goal of chapter 4 is to establish an equivalence result between the sufficient condition for idempotency provided by the FIC in (37) and the FTI just stated. This equivalence thus provides an intuitive metric interpretation of the apparently technical FIC: the equivalence says that the FIC effectively simply requires the faithfulness constraints to measure phonological distance in a sensible way, namely in compliance with the triangle inequality. The fact that all basic faithfulness constraints satisfy the FIC thus simply reflects the fact that phonologists have posited faithfulness constraints with good metrical properties.

### 2.5. Idempotency and the triangle inequality: the HG case (section 4.2)

In order to prepare the ground for the equivalence result between the FIC and the metric FTI, it useful to make a detour from OT to the related framework of *Harmonic Grammar* (HG; Legendre et al. 1990b,a; Smolensky and Legendre 2006), which allows for a more pristine view of the role played by the triangle inequality in the theory of idempotency. Section 4.2 thus repeats for HG the same reasoning initially developed in section 3.2 for OT. The result of this reasoning is that the HG grammar corresponding to any weighting of a given constraint set is idempotent provided each faithfulness constraint satisfies the implication (40). This implication is thus referred to as the *HG faithfulness idempotency condition* ( $\text{FIC}^{\text{HG}}$ ; from now on, the FIC obtained above for OT is referred to as the  $\text{FIC}^{\text{OT}}$ ).

$$(40) \quad \begin{array}{l} \text{For every choice of the threshold } \xi \geq 0: \\ \text{If: } \quad F(\mathbf{b}, \mathbf{c}) \leq \xi \\ \text{Then: } F(\mathbf{a}, \mathbf{c}) \leq F(\mathbf{a}, \mathbf{b}) + \xi \end{array}$$

The  $\text{FIC}^{\text{OT}}$  (37) is a special case of the  $\text{FIC}^{\text{HG}}$  (40) corresponding to the choice  $\xi = 0$ . The fact that the  $\text{FIC}^{\text{HG}}$  is stronger than the  $\text{FIC}^{\text{OT}}$  makes good sense: HG typologies properly contain OT typologies (at least when constraint violations are bounded), so that a stronger condition is expected to be needed in order to discipline a larger typology of grammars to all comply with idempotency. Crucially, the  $\text{FIC}^{\text{HG}}$  (40) is trivially equivalent to the metric FTI (107). This equivalence holds under no additional assumptions on the faithfulness constraints, the phonological candidates, or their correspondence relations. This equivalence says that idempotency in HG is tightly linked to the metric properties of the faithfulness constraints, as the sufficient condition for HG idempotency provided by the  $\text{FIC}^{\text{HG}}$  simply requires the faithfulness constraints to measure phonological distances in a sensible way, namely in compliance with the metric triangle inequality.

### 2.6. Faithfulness additivity, categoricity, and monotonicity (section 4.3)

The sufficient condition for OT idempotency provided by the  $\text{FIC}^{\text{OT}}$  turns out not to be equivalent to the metric FTI in the general case. This is expected, given that the FTI has been found to be equivalent to the  $\text{FIC}^{\text{HG}}$  which is in turn stronger than the  $\text{FIC}^{\text{OT}}$ . Although the equivalence between the  $\text{FIC}^{\text{OT}}$  and the FTI fails in the general case, one might expect the equivalence to hold for special (hopefully realistic) classes of faithfulness constraints. To this end, section 4.3 looks at McCarthy's (2003b) conjecture that all constraints relevant for phonological theory are *categorical*. He provides an explicit formalization of this intuition for markedness constraints (see his scheme (1) on p. 77). His treatment of faithfulness constraints is not as explicit: he discusses individual faithfulness constraints but does not provide a general scheme. Section 4.3 fills this gap: it carefully formulates the notion of faithfulness categoricity and introduces the two related notions of faithfulness additivity and monotonicity.

Here is an informal preview. Intuitively, a faithfulness constraint  $F$  is *additive* whenever a phonological candidate can be broken up into smaller “sub-candidates” in such a way that the number of violations assigned by  $F$  to the candidate is the sum of the number of violations assigned to the “sub-candidates”, as informally stated in (41).

$$(41) \quad F(\text{candidate}) = \sum_{\text{sub-candidates}} F(\text{sub-candidate})$$

An additive faithfulness constraint  $F$  is called *categorical* provided the sub-candidates can be chosen “small enough” that they are assigned either 0 or 1 violations, but never more than 1, as stated in (42). The notion of “sub-candidate” thus formalizes McCarthy’s intuitive notion of *locus of violations*.

$$(42) \quad F(\text{sub-candidate}) = 0 \text{ or } 1$$

A faithfulness constraint is *monotone* provided the number of violations grows when the candidate gets “larger”, in the sense of having a larger number of sub-candidates, as stated in (43). Intuitively, additivity entails monotonicity: a larger candidate has more sub-candidates, yielding a sum (41) with more non-negative terms.

$$(43) \quad \text{candidate}_{\text{small}} \subseteq \text{candidate}_{\text{large}} \implies F(\text{candidate}_{\text{small}}) \leq F(\text{candidate}_{\text{large}})$$

According to *correspondence theory* (McCarthy and Prince 1995), a candidate is a triplet consisting of an underlying form, a surface form, and a correspondence relation between them. Hence, a candidate can be split into sub-candidates along three dimensions: the underlying string can be split into sub-strings; the surface string can be split into sub-strings; and the correspondence relations can be split into sub-relations. These informal considerations thus yield three notions of additivity, categoricity, and monotonicity, carefully formalized in section 4.3.

McCarthy’s original conjecture that all faithfulness constraints relevant for phonological theory are “categorical” can thus be formalized by requiring them to be categorical relative to the input, or relative to the output, or relative to the correspondence relation. A slight strengthening on this conjecture is considered and shown to hold for the faithfulness constraints used in the phonological literature. This strengthened conjecture requires all faithfulness constraints to be: categorical relative to the correspondence relation; or categorical relative to the underlying form and furthermore monotone relative to the surface form; or categorical relative to the surface form and furthermore monotone relative to the underlying form. This slight asymmetry between categoricity relative to the correspondence relation (which does not require a corresponding monotonicity assumption) and categoricity relative to the underlying or surface form (which instead does come with a corresponding monotonicity requirement) has to do with subtle differences in the formalization of the three notions of categoricity. I will refer to the latter strengthened conjecture as the *categoricity-plus-monotonicity conjecture*.

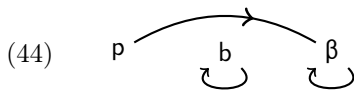
## 2.7. Idempotency and the triangle inequality: the OT case (section 4.4)

Section 3.2 has established the  $\text{FIC}^{\text{OT}}$  (37) as a sufficient condition for the idempotency of OT grammars. Section 4.1 has introduced the requirement that faithfulness constraints measure phonological distance in compliance with the triangle inequality, as captured by the FTI (107). What is the relationship between the  $\text{FIC}^{\text{OT}}$  and the FTI? The metric FTI entails the  $\text{FIC}^{\text{OT}}$  for any faithfulness constraint (indeed, the FTI is equivalent to the  $\text{FIC}^{\text{HG}}$  which in turn entails the  $\text{FIC}^{\text{OT}}$ ). The reverse entailment fails for general faithfulness constraints, as recalled above. Yet, is it possible to secure the reverse entailment from the  $\text{FIC}^{\text{OT}}$  to the FTI for special classes of faithfulness constraints? To start, I note that the entailment from the  $\text{FIC}^{\text{OT}}$  to the FTI trivially holds for faithfulness constraints which are *binary*, namely assign only zero or one violations. Obviously, binary faithfulness constraints are useless in phonology. Yet, this trivial observation concerning binary faithfulness constraints provides the starting point for a stronger result, which can be informally previewed as follows. Recall that a faithfulness constraint is additive provided the number of violations it assigns to a candidate is equal to the sum of the number of violations it assigns to the corresponding sub-candidates. The faithfulness constraint is furthermore categorical provided the sub-candidates can be chosen in such a way that they are assigned either zero or one violations. In other words, a categorical constraint is binary when restricted to the sub-candidates. Since the entailment from the  $\text{FIC}^{\text{OT}}$  to the FTI holds for binary faithfulness constraint, one then expects the entailment to extend to categorical faithfulness constraints, by summing over sub-candidates. Section 4.4 formalizes this intuition into the proof of an equivalence result between the FTI and the  $\text{FIC}^{\text{OT}}$  for

faithfulness constraints which satisfy the categoricity-plus-monotonicity conjecture stated above. The additional monotonicity requirement is technical grease needed by the non-trivial proof machinery. This equivalence result is the main technical result contained in this document.

### 2.8. Output-drivenness and counter-bleeding opacity (section 5.1)

Tesar (2013) investigates another structural condition on phonological grammars, which he calls *output-drivenness*. It formalizes the intuition that any discrepancy between an underlying and a surface (or *output*) form is *driven* exclusively by the goal of making the surface form fit the phonotactics. It thus captures the intuition that, if /p/ is mapped to [β], then /b/ should be mapped to [β] as well, as /b/ is more similar to [β] than /p/ is. In other words, output-drivenness rules out *derived environment effects* or *saltations* (Łubowicz 2002; White 2013) such as the one in (44). Output-drivenness is predicated on a notion of relative *similarity* (see below for more details). For instance, the phonological pattern in (44) only fails at output-drivenness under the assumption that /b/ is more similar to [β] than /p/ is (say, because the former only differs for voicing, while the latter differs for both voicing and continuancy).



Output-drivenness entails idempotency: if phonological forms are only repaired to satisfy the phonotactics (output-drivenness), phonotactically licit forms are faithfully realized (idempotency). The reverse entailment fails: output-drivenness is a stronger condition than idempotency. This is shown for instance by the saltation in (44), which fails at output-drivenness (under a plausible definition of the similarity order) and yet succeeds at idempotency (as the two phonotactically licit forms **b**, **β** are faithfully mapped to themselves). Within SPE, saltations are the result of *counter-bleeding* rule interactions (Baković 2011 and references therein). In conclusion, the notion of output-drivenness is relevant to phonology because its negation captures counter-feeding opacity (chain shifts) and counter-bleeding opacity (saltations).

### 2.9. Output-drivenness in OT (section 5.2)

Which assumptions on the constraint set suffice to guarantee that the OT grammar corresponding to any constraint ranking is output-driven? Tesar settles this question by showing that it suffices that each faithfulness constraint in the constraint set satisfies the two implications (170) for every two candidates (**a**, **d**) and (**b**, **d**) such that **a** is less similar to **d** than **b** is (according to the relevant notion of similarity), for every other candidate (**b**, **c**), and for some candidate (**a**, **c**). Condition (170) is thus referred to as the *faithfulness output-drivenness condition* (FODC).

(45) 

a.	<b>If:</b> $F(\mathbf{a}, \mathbf{d}) < F(\mathbf{a}, \mathbf{c})$
	<b>Then:</b> $F(\mathbf{b}, \mathbf{d}) < F(\mathbf{b}, \mathbf{c})$
b.	<b>If:</b> $F(\mathbf{b}, \mathbf{c}) < F(\mathbf{b}, \mathbf{d})$
	<b>Then:</b> $F(\mathbf{a}, \mathbf{c}) < F(\mathbf{a}, \mathbf{d})$

Section 5.2 reviews Tesar's theory of OT output-drivenness leading to a proof of the sufficiency of the FODC. The FODC are easily seen to entail the FIC (37), matching the fact that output-drivenness entails idempotency.

### 2.10. An axiomatic definition of the similarity order (sections 5.3-5.4)

As anticipated above, the notion of output-drivenness (as well as the sufficient condition for output-drivenness provided by the FODC) is predicated upon a notion of relative similarity. This notion of similarity is formalized through a partial order  $\leq_{\text{sim}}$  on the candidate set which intuitively orders two candidates (which share the same surface form) according to their degree of internal similarity. How should the partial order  $\leq_{\text{sim}}$  which induces similarity comparisons be defined? Section 5.3 reviews Tesar's (2013) specific definition of similarity. It is stated concretely in terms of strings and correspondence relations and it is tailored to a faithfulness constraint set which only consists of three types of faithfulness constraints, namely MAX, DEP, and IDENT. Tesar indeed establishes that these three types of faithfulness constraints satisfy the FODC (in the special case where all correspondence relations are one-to-one). The latter result, although quite limited, turns out to be non-trivial. For instance, two pages of Tesar's book suffice to

establish the FODC as a sufficient condition for OT output-drivenness, while the entire chapter 3 is devoted to verifying the FODC for just these three constraint types MAX, DEP, and IDENT.

Section 5.4 thus explores an alternative approach. I start from the intuition that, within constraint-based phonology, it is natural to assess similarity through the faithfulness constraints. Thus, instead than a specific, concrete definition of the similarity order, I consider a general axiom on the similarity order, stated in terms of faithfulness constraints. The proposed axiom is parameterized by an arbitrary faithfulness constraint set  $\mathcal{F}$  (which might be smaller than or different from the faithfulness constraint set used to define the typology). The proposed axiom roughly requires the less similar candidate to violate each faithfulness constraint in  $\mathcal{F}$  more than the more similar candidate. The reasoning in Tesar (2013; chapter 3) can be re-booted to show that Tesar’s concrete definition of the similarity order satisfies the proposed axiom as a special case (when the faithfulness constraint set  $\mathcal{F}$  consists of MAX, DEP, and IDENT). His theory of output-drivenness thus counts as a special case of the theory of output-drivenness developed for the axiomatized similarity order.

### 2.11. Output-drivenness and the triangle inequality (section 5.5)

The main result of section 5.5 is that the theory of output-drivenness relative to the axiomatized similarity order turns out to be related to the requirement that faithfulness constraints measure phonological distance in compliance with the metric triangle inequality, as captured by the FTI (107). In particular, I show that the FTI entails Tesar’s FODC relative to the axiomatized notion of similarity order. This entailment holds for any faithfulness constraint which belongs to the selected faithfulness subset  $\mathcal{F}$  used to measure similarity. Furthermore, this entailment holds under no assumptions on the faithfulness constraints (such as the categoricity-plus-monotonicity restriction). What about the reverse entailment from the FODC to the metric FTI? As recalled above, the  $\text{FODC}^{\text{OT}}$  entails the sufficient condition for idempotency provided by the  $\text{FIC}^{\text{OT}}$ . Furthermore, we have seen that the  $\text{FIC}^{\text{OT}}$  entails the metric FTI for those faithfulness constraints which satisfy the categoricity-plus-monotonicity conjecture. I conclude that the  $\text{FODC}^{\text{OT}}$ , the  $\text{FIC}^{\text{OT}}$  and the FTI are equivalent for any categoricity-plus-monotone faithfulness constraint (which furthermore belongs to the faithfulness constraint subset  $\mathcal{F}$  used in the axiomatic definition of the similarity order). This equivalence affords a substantial simplification of Tesar’s theory. In fact, it says that all constraints that have been established in sections 3.3-3.5 to satisfy the  $\text{FIC}^{\text{OT}}$ , actually also satisfy the FODC (as they are all categoricity-plus-monotone). A large set of faithfulness constraints (beyond the three considered by Tesar) are thus shown in a snap to satisfy the FODC.

### 2.12. Output-drivenness in HG (section 5.6)

In the case of idempotency, we have seen above that the connection with the triangle inequality was enhanced by switching from OT to HG. The situation is analogous for the case of output-drivenness. Section 5.6 adapts Tesar’s reasoning from OT to HG. The result of this reasoning is that the HG grammar corresponding to any weighting of a given constraint set is output-driven relative to an arbitrary similarity order  $\leq_{\text{sim}}$  provided each faithfulness constraint satisfies the implication (46) for every two candidates  $(\mathbf{a}, \mathbf{d})$  and  $(\mathbf{b}, \mathbf{d})$  such that  $(\mathbf{a}, \mathbf{d}) \leq_{\text{sim}} (\mathbf{d}, \mathbf{b})$ , for every other candidate  $(\mathbf{b}, \mathbf{c})$ , and for some candidate  $(\mathbf{a}, \mathbf{c})$ . Condition (46) is thus referred to as the HG *faithfulness output-drivenness condition* ( $\text{FODC}^{\text{HG}}$ ; from now on, the FODC obtained above for OT is referred to as the  $\text{FODC}^{\text{OT}}$ ).

$$(46) \quad \boxed{\begin{array}{l} \text{For every choice of the threshold } \xi: \\ \text{If: } \quad F(\mathbf{b}, \mathbf{c}) \leq F(\mathbf{b}, \mathbf{d}) + \xi \\ \text{Then: } F(\mathbf{a}, \mathbf{c}) \leq F(\mathbf{a}, \mathbf{d}) + \xi \end{array}}$$

It can be shown that the  $\text{FODC}^{\text{OT}}$  (170) is a special case of the  $\text{FODC}^{\text{HG}}$  (46) corresponding to the choice  $-1 < \xi < +1$ . The fact that the  $\text{FODC}^{\text{HG}}$  is stronger than the  $\text{FODC}^{\text{OT}}$  makes good sense: HG typologies properly contain OT typologies (at least when constraint violations are bounded), so that a stronger condition is expected to be needed in order to discipline a larger typology of grammars to all comply with output-drivenness. Crucially, the  $\text{FODC}^{\text{HG}}$  (46) is trivially equivalent to the metric FTI (107). This equivalence holds under no additional assumptions on the faithfulness constraints, the phonological candidates, or their correspondence relations.

### 2.13. A network of relationships among faithfulness conditions (section 6.1)

Let me take stock. Idempotency requires any phonotactically licit form to be faithfully realized. Output-drivenness requires any discrepancy between underlying and surface (or output) forms to be driven by the phonotactics. Chapters 3-5 have systematized the theory of idempotency and Tesar’s (2013) theory of OT output-drivenness within OT and HG. These theories provide tight guarantees for OT and HG idempotency and output-drivenness through conditions on the faithfulness constraints, referred to as the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FIC}_{\text{comp}}^{\text{HG}}$ , the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  and the  $\text{FODC}_{\text{comp}}^{\text{HG}}$ . These four conditions are technical conditions which lack *prima facie* any intuitive interpretation. To provide such an interpretation, I have dug deeper into the formal underpinning of the theory of faithfulness. Intuitively, faithfulness constraints measure the phonological distance between underlying and surface forms. It thus makes sense to investigate whether faithfulness constraints satisfy axioms of the abstract notion of distance. I have focused on one such axiom, the *triangle inequality*. The main result obtained is that:

- (47) The four abstract conditions  $\text{FIC}^{\text{OT}}$ ,  $\text{FODC}^{\text{OT}}$ ,  $\text{FIC}^{\text{HG}}$  and  $\text{FODC}^{\text{HG}}$  for idempotency and output-drivenness in OT and HG are all equivalent to the faithfulness triangle inequality.

The four conditions can thus all be interpreted as simply requiring the faithfulness constraints to measure phonological distance in compliance with a core axiom of the abstract notion of distance. This equivalence (47) holds for categorical faithfulness constraints in the case of the two OT conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and  $\text{FODC}_{\text{comp}}^{\text{OT}}$ , while it holds without restrictions on the nature of the faithfulness constraints in the case of the two HG conditions  $\text{FODC}_{\text{comp}}^{\text{HG}}$  and  $\text{FIC}_{\text{comp}}^{\text{HG}}$ .

Section 6.1 discusses various implications of this metric interpretation (47) of the four conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FODC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FIC}_{\text{comp}}^{\text{HG}}$ , and  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for idempotency and output-drivenness. First, (47) entails that idempotency and output-drivenness do not require stronger constraint conditions in HG than in OT, at least when we restrict ourselves to categorical constraints, as independently conjectured by McCarthy (2003b). Second, (47) entails that the conditions for idempotency and output-drivenness (in either OT or HG) are equivalent when we restrict ourselves to categorical constraints, under the additional assumption that the similarity order which underlies output-drivenness takes into account all of the faithfulness constraints in the constraint set used to define the typology. In other words, the only way to obtain idempotent grammars which fail at output-drivenness is to define output-drivenness relative to a similarity order which is blind to some of the faithfulness constraints. Third, (47) allows the results on which constraints satisfy the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  obtained in sections 3.3-3.5 to be extended in a snap to the other three conditions  $\text{FODC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FIC}_{\text{comp}}^{\text{HG}}$ , and  $\text{FODC}_{\text{comp}}^{\text{HG}}$ .

### 2.14. Opacity between output-drivenness and input-drivenness (sections 6.2 and 6.3)

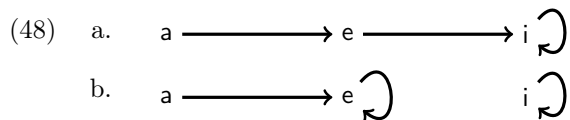
As anticipated, idempotency and output-drivenness are related to phonological opacity. Indeed, section 6.3 recalls that counter-feeding opacity is equivalent (under mild assumptions) to chain shifts and thus corresponds to a non-idempotent grammar. Furthermore, counter-bleeding opacity is equivalent (under mild assumptions) to (a generalized notion of) saltations and thus corresponds to an idempotent but not output-driven grammar. I then probe deeper into the theory of opacity by comparing Tesar’s notion of output-drivenness with a closely related notion of *input-drivenness*. The intuition can be previewed as follows. Tesar’s notion of output-drivenness is based on an *interpretation* perspective (motivated by the goal of contributing to the foundation of algorithms for learning lexicons of underlying forms from a given paradigm of surface forms). For instance, Tesar’s notion of similarity order compares two candidates which share the same surface form and only differ for the underlying form. The alternative notion of input-drivenness instead adopts a *production* perspective. For instance, it is based on a notion of similarity order which compares two candidates which share the same underlying form and only differ for the surface form. Like Tesar’s original notion, also this alternative notion of input-drivenness fails for chain shifts and saltations, so that its negation captures both counter-feeding opacity (chain shifts) and counter-bleeding opacity (saltations). Furthermore, like Tesar’s original notion, also this alternative notion of output-drivenness turns out to be related to the condition that the faithfulness constraints measure phonological distance in compliance with the metric triangle inequality. I conclude with some remarks on how to pull apart Tesar’s original notion of output-drivenness and the alternative notion of input-drivenness.

### 2.15. Idempotency and chain shifts (section 6.4)

The FIC is only a *sufficient* condition for OT idempotency. Yet, for any faithfulness constraint which fails at the FIC, it is easy to construct a simple case where idempotency fails, leading to a chain shift such as the pattern  $a \rightarrow e \rightarrow i$  in (34). Any analysis of chain shifts within (classical) OT requires at least one faithfulness constraint which does not satisfy the FIC. From this perspective, the results obtained in sections 3.2-3.5 concerning which faithfulness constraints happen to satisfy the FIC under which conditions can be reinterpreted as a toolkit for modeling chain shifts. Articulating the phonological implications of the formal theory of idempotency, section 6.4 systematizes various approaches to chain shifts in the classical OT literature by showing that they differ for how they choose the culprit faithfulness constraint from the list of non-FIC abiding constraints compiled in sections 3.3-3.5. Particular attention is paid to approaches to chain shifts which rely on restricted faithfulness constraints, which are shown to offer an exact translation of the classical rule-based counter-feeding approach to chain shifts.

### 2.16. Benign chain shifts (section 6.5)

Nine-month-olds already react differently to licit and illicit sound combinations (Jusczyk et al. 1993), thus displaying knowledge of the target adult phonotactics. The literature on this early acquisition of phonotactics usually assumes that the learner posits a fully faithful underlying form for each phonotactically licit training surface form (Gnanadesikan 2004; Hayes 2004; Prince and Tesar 2004). Is this assumption of faithful underlying forms computationally sound? or could it instead doom the learner to positing inconsistent mappings? For instance, if the target grammar is not idempotent and rather enforces a chain shift such as (48a), the assumption that the phonotactically licit  $e$  is faithfully mapped to itself might be dangerous.



Yet, the chain shift (48a) raises no issues for the learner's assumption of fully faithful underlying forms whenever the chain shift happens to be *benign*, in the sense that the typology entertained by the learner happens to contain another grammar such as (48b), which is idempotent (there is no chain shift) and yet makes the same phonotactic distinctions ( $[a]$  is illicit and  $[e, i]$  licit for both grammars). The theory of idempotency in classical OT systematized in section 6.4 says that there are just three strategies to derive chain shifts in OT: through the conjunction of two faithfulness constraints which individually satisfy the FIC; through the restriction of a faithfulness constraint whose unrestricted counterpart satisfies the FIC; and through breaking of underlying forms into multiple surface forms. Are the chain shifts obtained through these three means benign? Section 6.5 offers some preliminary discussion on this question, towards a computational foundation for a variety of models of the acquisition of phonotactics which share the assumption of completely faithful underlying forms.

## The theory of idempotency in Optimality Theory

This chapter contributes to a research program in constraint-based phonology which aims at distilling analytically the implications of constraint theories for formal typological properties (Prince 2007). For instance, Moreton (2004b) develops constraint conditions for the property of *eventual idempotency* and Tesar (2013) develops constraint conditions for the property of *output-drivenness* (which I'll discuss in chapter 5). This chapter focuses on a third formal property which is intermediate between those two, namely *idempotency*. Building on Tesar's analysis of output-drivenness, this chapter develops tight sufficient conditions for idempotency within classical *Optimality Theory* (OT; Prince and Smolensky 2004; Moreton 2004b). Chapter 4 will then extend the theory of idempotency to the case of *Harmonic Grammar* (Legendre et al. 1990b,a; Smolensky and Legendre 2006) and discuss the relationship between idempotency and Tesar's output-drivenness.

A formal theory of idempotency is relevant both for phonological theory and for modeling the acquisition of phonology. In fact, idempotency is related to opacity: a grammar fails at idempotency provided it displays a *chain shift*, which corresponds to counter-feeding ordering in a rule-based phonological framework. Understanding the conditions which ensure idempotency thus yields a toolkit for modeling chain shifts, which have proven recalcitrant to constraint-based analyses. Furthermore, various models of the early acquisition of phonotactics (Gnanadesikan 2004; Hayes 2004; Prince and Tesar 2004) assume that the learner posits a fully faithful underlying form for each training phonotactically licit surface form. These models thus effectively assume that the typology explored by the learner consists of idempotent grammars. Chapter 6 will explore the implications of the analysis of idempotency developed here for modeling chain shifts and for modeling the early acquisition of phonotactics.

This chapter is organized as follows. Section 3.1 formalizes the intuition that a phonological grammar is *idempotent* provided it maps phonotactically licit forms faithfully to themselves. Equivalently, provided it displays no chain shifts. The rest of the chapter then establishes tight sufficient conditions for idempotency in (classical) Optimality Theory. Building on Tesar (2013), these conditions are derived in two steps. First, section 3.2 shows that idempotency follows from a formal condition on the faithfulness constraints, called the *faithfulness idempotency condition* (FIC). Second, sections 3.3, 3.4, and 3.5 shown that this FIC is satisfied by a variety of faithfulness constraints which naturally arise within McCarthy and Prince's (1995) Correspondence Theory of faithfulness. Chapter 4 will provide an intuitive interpretation of the FIC based on the intuition that faithfulness constraints measure the phonological distance between underlying and surface forms.

### 3.1. Idempotency

This section introduces the notion of idempotent phonological grammar within a representational framework which is a segmental version of McCarthy and Prince's (1995) Correspondence Theory.

#### 3.1.1. Representational framework

Consider a finite set of *segments* (for instance, the segments in the IPA table, or some subset thereof), denoted by  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$ . Strings obtained through segment concatenation are denoted by  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$ . The notation  $\mathbf{a} = \mathbf{a}_1 \cdots \mathbf{a}_\ell$  says that the string  $\mathbf{a}$  is the concatenation of the segments  $\mathbf{a}_1, \dots, \mathbf{a}_\ell$  and thus has *length*  $\ell$ . Throughout this document, I assume the representational framework (49). Underlying and surface forms are strings of segments. Phonological candidates establish a correspondence between the segments of these underlying and surface strings.



- (49) The *candidate set* consists of triplets  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  of an *underlying* segment string  $\mathbf{a}$  and a *surface* segment string  $\mathbf{b}$  together with a *correspondence relation*  $\rho_{\mathbf{a}, \mathbf{b}}$  between the segments of  $\mathbf{a}$  and those of  $\mathbf{b}$ .<sup>1</sup>

Correspondence relations will be denoted by thin lines. To illustrate, (50) represents the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose underlying string  $\mathbf{a}$  is /bnɪk/, whose surface string  $\mathbf{b}$  is [blik], whose correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  maps underlying to surface segments respecting their position in the strings.

$$(50) \quad \begin{array}{l} \mathbf{a} = \text{bn} \text{ɪ} \text{k} \\ \quad \quad \quad | | | | \\ \mathbf{b} = \text{bl} \text{ɪ} \text{k} \end{array}$$

The representational assumption (49) places no *a priori* restrictions neither on the underlying and surface strings nor on the relations which put them in correspondence. The representational framework is thus sufficiently flexible to encompass approaches which hardwire some (universal) restrictions into the candidate set (Blaho et al. 2007). This flexibility will be exploited in the rest of this chapter, which will explore the implications of various restrictions on the correspondence relations which can figure in the candidate set.

### 3.1.2. Identity candidates

Idempotency is about phonotactically licit forms being mapped to themselves. It thus requires the distinction between underlying and surface forms to be blurred. This is achieved through axiom (51). It can be interpreted as a candidacy *reflexivity axiom*, as it requires each (surface) string to be in correspondence with itself. This axiom will play a crucial role in the definition of idempotency in the next subsection.

- (51) If the candidate set contains a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  with a surface form  $\mathbf{b}$ , it also contains the corresponding *identity* candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , where  $\mathbb{I}_{\mathbf{b}, \mathbf{b}}$  is the identity correspondence relation among the segments of  $\mathbf{b}$ .

By (51), any surface form can be construed as an underlying form (of the corresponding identity candidate). In other words, the set of surface forms is a subset of the set of underlying forms. This is a slightly weaker condition than Moreton’s (2004b) *homogeneity*, which requires the sets of underlying and surface forms to coincide. Both reflexivity and homogeneity hold when underlying and surface representations are constructed out of the same “building blocks”. Moreton claims that “most phonological representations are in fact present in both [underlying and surface forms]” so that reflexivity and homogeneity hold for “much of the core business of phonology.” Yet, the reflexivity axiom (51) obviously does not hold in full generality. To illustrate, suppose that the candidate set contains the candidate  $(\mathbf{a}, \mathbf{b}) = (/mabap/, [ma.ba])$ . The reflexivity axiom (51) requires the candidate set to also contain the candidate  $(\mathbf{b}, \mathbf{b}) = (/ma.ba/, [ma.ba])$ . This contravenes the plausible assumption that syllabification is a property of the surface representations and is absent in the underlying representations. In the case of constraint-based phonology, this difficulty can be circumvented by switching from the identity candidates required by (51) to McCarthy’s (2002) *fully faithful candidates*, as explained in subsection 3.2.5.

### 3.1.3. Idempotency

Within the representational framework just defined, a phonological grammar is a map  $G$  which takes an underlying form  $\mathbf{a}$  and returns a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose underlying string is indeed  $\mathbf{a}$ .<sup>2</sup> A string  $\mathbf{b}$  is called *phonotactically licit* according to a grammar  $G$  provided there exists at least one string  $\mathbf{a}$  (with  $\mathbf{a}$  possibly identical to  $\mathbf{b}$ ) such that the grammar  $G$  maps the underlying form  $\mathbf{a}$  to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose surface string is  $\mathbf{b}$ . A grammar  $G$  is *idempotent* provided it maps any phonotactically licit surface form to itself, as formalized by the implication (52) in the following definition. The antecedent of the implication says that the surface form  $\mathbf{b}$  is phonotactically licit relative to the grammar  $G$ , because it is

<sup>1</sup> Correspondence relations might want to distinguish between multiple occurrences of the same segment in a string. Thus, correspondence relations cannot be defined simply as relations between the two *sets* of underlying and surface segments. To keep the presentation straightforward, I follow common practice and ignore these subtleties.

<sup>2</sup> For the sake of simplicity, I assume that a grammar maps an underlying form to a *single* candidate. This assumption is not crucial and the results obtained extend to a framework where grammars map an underlying form to a *set* of candidates, thus modeling phonological variation.

the surface realization of some underlying form  $\mathbf{a}$ . The consequent says that  $\mathbf{b}$  is then mapped faithfully to itself. The reflexivity axiom (51) ensures the existence of the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  in the consequent of (52).

**Definition 1.** [Idempotency] A grammar  $G$  is *idempotent* provided it satisfies the implication

$$(52) \quad \begin{array}{l} \mathbf{If:} \quad G(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) \\ \mathbf{Then:} \quad G(\mathbf{b}) = (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}}) \end{array}$$

for any candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in the candidate set.  $\square$

To illustrate, suppose that a grammar raises the low vowel /a/ to [e]. The mid vowel [e] is therefore phonotactically licit. In order for that grammar to comply with condition (52) and thus qualify as idempotent, the underlying form /e/ must be mapped faithfully to [e].<sup>3</sup>

### 3.1.4. Chain shifts

The candidate set might provide many different relations  $\rho_{\mathbf{b}, \mathbf{b}}$  to put a string  $\mathbf{b}$  in correspondence with itself. Some options are illustrated in (53) in the case where  $\mathbf{b} = \mathbf{amba}$ . Axiom (51) requires one of these correspondence relations provided by the candidate set to be the identity relation  $\mathbb{I}_{\mathbf{b}, \mathbf{b}}$ , illustrated the left-most candidate in (53). This identity relation  $\mathbb{I}_{\mathbf{b}, \mathbf{b}}$  is intuitively the *best* way to put the string  $\mathbf{b}$  in correspondence with itself. A grammar  $G$  is *well-behaved* provided it abides to this intuition: whenever  $G$  maps an underlying string  $\mathbf{b}$  to the same surface string  $\mathbf{b}$ , it does so through the identity correspondence relation  $\mathbb{I}_{\mathbf{b}, \mathbf{b}}$ . In other words  $G(\mathbf{b}) = (\mathbf{b}, \mathbf{b}, \rho_{\mathbf{b}, \mathbf{b}})$  is impossible when  $\rho_{\mathbf{b}, \mathbf{b}} \neq \mathbb{I}_{\mathbf{b}, \mathbf{b}}$ .

$$(53) \quad \begin{array}{ccc} \mathbf{b} = \mathbf{a m b a} & \mathbf{b} = \mathbf{a m b a} & \mathbf{b} = \mathbf{a m b a} \\ \quad \quad \quad \begin{array}{c} | \quad | \quad | \quad | \\ | \quad | \quad | \quad | \end{array} & \quad \quad \quad \begin{array}{c} | \quad \times \quad | \\ | \quad \times \quad | \end{array} & \quad \quad \quad \begin{array}{c} \vee \quad | \quad | \\ \vee \quad | \quad | \end{array} \\ \mathbf{b} = \mathbf{a m b a} & \mathbf{b} = \mathbf{a m b a} & \mathbf{b} = \mathbf{a m b a} \end{array}$$

Suppose now that a grammar  $G$  fails at the idempotency implication (52) for some candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , as stated in (54):  $G$  maps the underlying form  $\mathbf{a}$  to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , as required by the antecedent of the idempotency implication; but  $G$  fails to map the underlying form  $\mathbf{b}$  to the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , as required by the consequent.

(54) A grammar  $G$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  iff:

- a.  $G(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ ;
- b.  $G(\mathbf{b}) \neq (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ .

Condition (54b) means that  $G$  maps the underlying form  $\mathbf{b}$  to some candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ . This means that either the two strings  $\mathbf{b}$  and  $\mathbf{c}$  differ or  $\mathbf{b}$  and  $\mathbf{c}$  coincide but the two correspondence relations  $\rho_{\mathbf{b}, \mathbf{c}}$  and  $\mathbb{I}_{\mathbf{b}, \mathbf{b}}$  differ. The latter option is impossible when  $G$  is well-behaved. The strings  $\mathbf{b}$  and  $\mathbf{c}$  must thus differ and condition (54) becomes (55).

(55) A (well-behaved) grammar  $G$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  iff there is a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  such that:

- a.  $G(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ ;
- b.  $G(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ .

Condition (55) says that  $G$  maps  $\mathbf{a}$  to  $\mathbf{b}$  and then in turn maps  $\mathbf{b}$  to  $\mathbf{c}$ . Since  $\mathbf{b} \neq \mathbf{c}$ , this scheme  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  is called a *chain shift* in the phonological literature (see Łubowicz 2011 for a comprehensive review). In conclusion, a (well-behaved) grammar  $G$  fails at idempotency if and only if it enforces chain shifts.

<sup>3</sup> Usually, idempotency is a notion which applies to a function  $f$  between a set  $X$  and itself and requires the identity  $f(f(x)) = f(x)$  for every argument  $x \in X$ . The connection between this notion and definition above is straightforward. Given a grammar  $G$ , let  $g$  be the corresponding string function, namely the function from strings to strings defined by the condition  $g(\mathbf{a}) = \mathbf{b}$  provided  $G(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  for some correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$ . The grammar  $G$  is idempotent according to the definition above if and only if the corresponding string function  $g$  satisfies the condition  $g(g(\mathbf{a})) = g(\mathbf{a})$  for any underlying string  $\mathbf{a}$ .

### 3.2. Idempotency in Optimality Theory

The notion of idempotency introduced in the preceding section is independent of any specific phonological framework used to define the grammar  $G$ . Starting with this section, I focus on the specific framework of *Optimality Theory* (OT; Prince and Smolensky 2004) (Magri to appear extends the theory of idempotency to the related framework of *Harmonic Grammar*). Which conditions guarantee that all the grammars in an OT typology are idempotent? The answer developed in this section has two parts: a condition on the candidate set, in the form of a candidacy *transitivity axiom*; and a condition on the violation profiles of the faithfulness constraints.

#### 3.2.1. Classical Optimality Theory (OT)

A *constraint*  $C$  is a function which takes a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and returns a *number of violations*  $C(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  which is large when the candidate scores poorly from the perspective relevant to that constraint. A constraint  $C$  *prefers* a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  to another candidate  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c}, \mathbf{d}})$  provided it assigns less violations to the former than to the latter, namely  $C(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) < C(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c}, \mathbf{d}})$ . A *constraint ranking* is an arbitrary linear order  $\gg$  over a set of constraints. A constraint ranking  $\gg$  *prefers* a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  to another candidate  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c}, \mathbf{d}})$  provided the  $\gg$ -highest constraint which assigns a different number of violations to the two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c}, \mathbf{d}})$ , prefers the former candidate to the latter. The OT grammar  $G_{\gg}$  corresponding to a ranking  $\gg$  maps an underlying form  $\mathbf{a}$  to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  which is preferred by the ranking  $\gg$  to all other candidates  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  which share that underlying form  $\mathbf{a}$ .<sup>4</sup>

A *faithfulness* constraint  $F$  has the property that it never assigns any violations to any identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , as stated in (56).

$$(56) \quad F(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}}) = 0$$

A *markedness* constraint  $M$  has the property that it is blind to underlying forms, so that it assigns the same number of violations to any two candidates  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  sharing the surface form  $\mathbf{c}$  (independently of their underlying forms), as stated in (57).

$$(57) \quad M(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) = M(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$$

Given the candidacy reflexivity axiom (51), no (non-trivial) constraint can be *both* a faithfulness and a markedness constraint. In fact, suppose by contradiction that were the case for some constraint  $C$ . Consider an arbitrary candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in the candidate set. The reflexivity axiom thus ensures that the candidate set also contains the corresponding identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ . These two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  share the surface form  $\mathbf{b}$ . Since  $C$  is a markedness constraint,  $C$  must assign the same number of violations to those two candidates, as stated in (58a). Since  $C$  is also a faithfulness constraint, it does not penalize the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , as stated in (58b).

$$(58) \quad C(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) \stackrel{(a)}{=} C(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}}) \stackrel{(b)}{=} 0$$

In conclusion,  $C$  does not penalize any candidate, and it is therefore trivial.<sup>5</sup>

Although no constraint can be *both* a faithfulness and a markedness constraint, it can easily be *neither* (for instance, *comparative markedness constraints* are neither; see McCarthy 2002, 2003a as well as section 6.4 below for additional references). To rule out the latter case, I assume that the constraint set only consists of faithfulness and markedness constraints (this is Moreton's 2004 *conservativity* assumption).

$$(59) \quad \text{constraint set} = \text{faithfulness constraints} \cup \text{markedness constraints}$$

Let me call *classical* the version of OT endowed with the latter restriction (59) on the constraint set.

<sup>4</sup> As noted in footnote 2, I assume that grammars map an underlying form to a single candidate. This condition holds for OT grammars provided the constraint set is sufficiently rich relative to the candidate set, in the following sense: for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  which share the underlying form  $\mathbf{a}$ , the constraint set contains a constraint  $C$  which assigns them a different number of violations.

<sup>5</sup> This conclusion crucially rests on the candidacy reflexivity axiom (51), which intuitively ensures that the candidate set has enough identity candidates. Without this axiom, the assumption that  $C$  is a faithfulness constraint would indeed have no bite, as the faithfulness definitional condition (56) is stated in terms of identity candidates.

### 3.2.2. A sufficient condition for chain shifts

The classical assumption (59) that each constraint is either a faithfulness constraint (56) or a markedness constraint (57) ensures that the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  harmonically bounds any candidate  $(\mathbf{b}, \mathbf{b}, \rho_{\mathbf{b}, \mathbf{b}})$  with  $\rho_{\mathbf{b}, \mathbf{b}} \neq \mathbb{I}_{\mathbf{b}, \mathbf{b}}$ . To illustrate, the left-most candidate in (53) outperforms the other candidates listed. In fact, faithfulness constraints cannot prefer  $(\mathbf{b}, \mathbf{b}, \rho_{\mathbf{b}, \mathbf{b}})$  to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , by (56); and markedness constraints do not distinguish between two such candidates, by (57). In other words, the OT grammar  $G_{\gg}$  corresponding to any ranking  $\gg$  is well-behaved in the sense of subsection 3.1.4:  $G_{\gg}(\mathbf{b}) \neq (\mathbf{b}, \mathbf{b}, \rho_{\mathbf{b}, \mathbf{b}})$  whenever  $\rho_{\mathbf{b}, \mathbf{b}} \neq \mathbb{I}_{\mathbf{b}, \mathbf{b}}$ . The characterization of non-idempotency in terms of chain shifts in subsection 3.1.4 thus applies to (classical) OT grammars. To distill the implications of that characterization, let me weaken the “if-and-only-if” statement (55) into the “if” statement (60). In fact, if the grammar  $G_{\gg}$  maps the underlying form  $\mathbf{a}$  to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  as stated in (55a), the ranking  $\gg$  must in particular prefer the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  to any other loser candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}})$ , as stated in (60a). Furthermore, if the grammar  $G_{\gg}$  maps the underlying form  $\mathbf{b}$  to the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  as stated in (55a), the ranking  $\gg$  must in particular prefer this candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  to the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , as stated in (60b).

- (60) If  $G_{\gg}$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , there exists some candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  such that:
- a.  $\gg$  prefers  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}})$ ,  
for any correspondence  $\rho_{\mathbf{a}, \mathbf{b}}$  in the candidate set (if any);
  - b.  $\gg$  prefers  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ .

Condition (60b) that the ranking  $\gg$  prefers  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  to the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  means that the constraint set contains a constraint which prefers  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  such that all the constraints which are ranked by  $\gg$  above it assign the same number of violations to the two candidates. By (56), this constraint which prefers  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  cannot be a faithfulness constraint and must instead be a markedness constraint  $M$ . Condition (60b) can thus be explicitated as (61b) and (61c).

- (61) If  $G_{\gg}$  fails at idempotency on  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , there exist a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  and a markedness constraint  $M$  such that:
- a.  $\gg$  prefers  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}})$ ,  
for any correspondence  $\rho_{\mathbf{a}, \mathbf{b}}$  in the candidate set (if any);
  - b.  $M$  assigns fewer violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  than to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ ;
  - c. any faithfulness or markedness constraint  $\gg$ -ranked above  $M$  assigns  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  and  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  the same number of violations.

Since faithfulness constraints assign no violations to identity candidates by (56), condition (61c) that any faithfulness constraint ranked above  $M$  assigns the same number of violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  and  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$  means that it assigns no violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ . Condition (61c) can thus be made explicit as in (62c) and (62d).

- (62) If  $G_{\gg}$  fails at idempotency on  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , there exist a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  and a markedness constraint  $M$  such that:
- a.  $\gg$  prefers  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}})$ ,  
for any relation  $\rho_{\mathbf{a}, \mathbf{b}}$  in the candidate set (if any);
  - b.  $M$  assigns fewer violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  than to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ ;
  - c. any faithfulness constraint  $\gg$ -ranked above  $M$  assigns no violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ ;
  - d. any markedness constraint  $\gg$ -ranked above  $M$  assigns the same number of violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  and  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ .

The designated markedness constraint  $M$  prefers  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , by (62b). Furthermore, it is blind to the underlying forms, by (57). Hence,  $M$  also prefers  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . Assumption (62a) thus requires  $M$  to be ranked below some constraint with the opposite preference. The latter constraint cannot be a markedness constraint, because of (62d). It must therefore be a faithfulness constraint. Condition (62a) can thus be made explicit as in (63a).

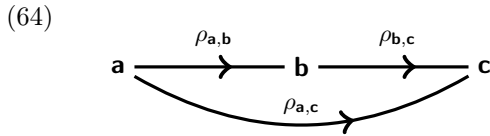
- (63) If  $G_{\gg}$  fails at idempotency on  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , there exist a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  and a markedness constraint  $M$  such that:

- a. for any correspondence  $\rho_{\mathbf{a},\mathbf{c}}$  in the candidate set (if any), there exists a faithfulness constraint  $\gg$ -ranked above  $M$  which assigns fewer violations to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  than to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ ;
- b.  $M$  assigns fewer violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  than to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ ;
- c. any faithfulness constraint  $\gg$ -ranked above  $M$  assigns no violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ ;
- d. any markedness constraint  $\gg$ -ranked above  $M$  assigns the same number of violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  and  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ .

Condition (63) just derived is necessary for idempotency to fail.

### 3.2.3. The faithfulness idempotency condition (FIC)

I am now ready to tackle the central question of this section: which conditions ensure that the the OT grammars corresponding to any ranking of a given constraint set is idempotent? The answer to this question is provided by the following proposition 1. The assumption made by the proposition is twofold. First, it restricts the candidate set: if it contains two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which share a string  $\mathbf{b}$  as the surface and underlying form respectively, it must also contain a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  which puts the underlying string  $\mathbf{a}$  of the former candidate in correspondence with the surface string  $\mathbf{c}$  of the latter candidate, as in (64).



Second, the assumption of the proposition restricts the constraint set: it requires all the faithfulness constraints to satisfy the implication (65), which is referred to as the *faithfulness idempotency condition* (FIC). The specific implication (36) in section 2.2 is a concrete example of the FIC.

**Proposition 1.** *Assume that, for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which share  $\mathbf{b}$  as the underlying and surface form respectively, the candidate set also contains a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  such that the following implication holds for every faithfulness constraint  $F$  in the constraint set.*

$$(65) \quad \text{If:} \quad F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0 \\ \text{Then:} \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$$

Then, the OT grammar corresponding to any ranking of the constraint set is idempotent.  $\square$

Proposition 1 follows straightforwardly from the discussion in the preceding subsection 3.2.2: the FIC (65) makes the two conditions (63a) and (63c) incompatible and thus prevents idempotency to fail. In fact, (63a) requires the designated markedness constraint  $M$  to be outranked by a faithfulness constraint  $F$  which assigns fewer violations to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  than to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ . This means that the consequent of the FIC (65) fails. The antecedent must therefore fail as well. This means in turn that  $F$  assigns some violations to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , contradicting (63c).

### 3.2.4. Composition candidates and the FIC<sub>comp</sub>

Proposition 1 makes no assumptions on the nature of the correspondence relation  $\rho_{\mathbf{a},\mathbf{c}}$  depicted in (64) and in particular on its relationship with the two other correspondence relations  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$ . For instance,  $\rho_{\mathbf{a},\mathbf{c}}$  could be the empty relation. This would make the FIC (65) trivial when  $F$  is an identity faithfulness constraint (because the quantity on the left-hand side of the inequality in the consequent would be equal to zero) but difficult when  $F$  is DEP or MAX (because the quantity on the left-hand side of the inequality would be large in this case). At the opposite extreme,  $\rho_{\mathbf{a},\mathbf{c}}$  could be the total relation, which puts any underlying segment in correspondence with any surface segment. This would make the FIC (65) trivial when  $F$  is DEP or MAX but difficult when  $F$  is an identity faithfulness constraint.

A natural assumption is that  $\rho_{\mathbf{a},\mathbf{c}}$  is the *composition*  $\rho_{\mathbf{a},\mathbf{c}} = \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  of the two correspondence relations  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$ .<sup>6</sup> This means that a segment  $\mathbf{a}$  of the string  $\mathbf{a}$  and a segment  $\mathbf{c}$  of the string  $\mathbf{c}$  are in correspondence through  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  if and only if there exists some “mediating” segment  $\mathbf{b}$  of the string  $\mathbf{b}$

<sup>6</sup> The operation of composition between two relations is usually denoted by “ $\circ$ ”. In the rest of this document, I write more succinctly  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  instead of  $\rho_{\mathbf{a},\mathbf{b}} \circ \rho_{\mathbf{b},\mathbf{c}}$ .

such that  $\mathbf{a}$  is in correspondence with  $\mathbf{b}$  through  $\rho_{\mathbf{a},\mathbf{b}}$  and furthermore  $\mathbf{b}$  is in correspondence with  $\mathbf{c}$  through  $\rho_{\mathbf{b},\mathbf{c}}$  (many examples will be provided in sections 3.3-3.5). The existence of this *composition* candidate is guaranteed by (66), which can thus be interpreted as a candidacy *transitivity axiom*, complementing the reflexivity axiom (51).

- (66) If the candidate set contains two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which share  $\mathbf{b}$  as the surface and the underlying form, it also contains the *composition candidate*  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  whose correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  is the composition of  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$ .

The original FIC (65) can now be specialized in terms of this composition candidate as the implication (67), which will be referred to as the FIC<sub>comp</sub> to highlight the fact that the left hand side of the inequality in the consequent features the composition candidate. The FIC<sub>comp</sub> entails the original FIC and thus provides a sufficient condition for the idempotency of all the grammars in an OT typology.

$$(67) \quad \text{If:} \quad F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0 \\ \quad \quad \quad \text{Then:} \quad F(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}_{\rho_{\mathbf{a},\mathbf{c}}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$$

The FIC<sub>comp</sub> is only a *sufficient* condition for idempotency, not a necessary-and-sufficient characterization of idempotency.<sup>7</sup> Yet, the FIC<sub>comp</sub> is a *tight* sufficient condition: for any faithfulness constraint which fails at the FIC<sub>comp</sub>, it is possible to construct a counterexample where idempotency indeed fails, as will be shown in section 6.4.

### 3.2.5. Refinements

The definition of idempotency in subsection 3.1.3 crucially relies on the existence of the identity candidate, as guaranteed by the reflexivity axiom (51). Yet, as discussed in subsection 3.1.2, this reflexivity axiom fails when surface representations are richer than underlying representations. For instance, the identity candidate  $(\mathbf{b}, \mathbf{b}) = (/ma.ba/, [ma.ba])$  makes no sense if syllabification is construed as a surface property. Within a constraint-based framework such as OT, this difficulty can be circumvented as follows.<sup>8</sup> Following McCarthy (2002, section 6.2),  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  is called a *fully faithful candidate* (FFC) relative to a constraint set provided it violates no faithfulness constraints in that constraint set. Identity candidates  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$  are FFCs, because of the definition (56) of faithfulness constraints. Yet, non-identity candidates can also qualify as FFCs. For instance, the candidate  $(/maba/, [ma.ba])$  is not the identity candidate and yet qualifies as a FFC, under the plausible assumption that syllabification of tautomorphic sequences is never contrastive and that no faithfulness constraint is therefore sensitive to syllabification. The reasoning presented in this section holds unchanged if idempotency is re-defined as follows: whenever  $G(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ , there exists a FFC  $(\beta, \mathbf{b}, \rho_{\beta,\mathbf{b}})$  such that  $G(\beta) = (\beta, \mathbf{b}, \rho_{\beta,\mathbf{b}})$ . This definition of idempotency does not require the existence of the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$  and thus dispenses with the problematic reflexivity axiom (51). Instead, it requires the following weaker axiom on the candidate set: for any candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  with a surface string  $\mathbf{b}$ , the candidate set also contains a FFC  $(\beta, \mathbf{b}, \rho_{\beta,\mathbf{b}})$  with that same surface form  $\mathbf{b}$ . This assumption complements McCarthy's (2002) assumption that each *underlying* form  $\mathbf{a}$  admits a FFC  $(\mathbf{a}, \alpha, \rho_{\mathbf{a},\alpha})$  with that underlying form.

## 3.3. Establishing the FIC<sub>comp</sub>: basic constraints

The preceding section has established the FIC<sub>comp</sub> (67) as a sufficient condition for idempotency in OT. This condition places no restrictions on the markedness constraints and instead only concerns the faithfulness constraints. The theory of idempotency in the rest of this chapter thus turns into an investigation of the formal underpinning of theories of faithfulness. Consider the *strictest* faithfulness constraint  $F_{\text{strictest}}$ , which is violated by every candidate but the identity candidate and thus demands perfect string identity. This constraint  $F_{\text{strictest}}$  satisfies the implication (68). In fact, the antecedent of (68) requires the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  to be the identity candidate. This means that the two strings  $\mathbf{b}$  and  $\mathbf{c}$  are identical and that the correspondence relation  $\rho_{\mathbf{b},\mathbf{c}}$  is the identity relation, so that the composition  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  coincides with  $\rho_{\mathbf{a},\mathbf{b}}$ . The equation in the consequent of (68) thus holds because the two candidates being compared are identical.

<sup>7</sup> Looseness has loomed at two steps in the derivation of the FIC<sub>comp</sub>. First, in the replacement of the if-and-only-if condition (55) with the if-condition (60). Second, in the replacement of the original FIC (65) for an arbitrary correspondence relation  $\rho_{\mathbf{a},\mathbf{c}}$  with the FIC<sub>comp</sub> (67) for the composition correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ .

<sup>8</sup> Thanks to an anonymous reviewer for discussion on the content of this subsection.







set) provided there is no underlying or surface string which contains a segment for which the feature  $\varphi$  is undefined. The identity faithfulness constraint  $\text{IDENT}_{\varphi}$  corresponding to a total feature  $\varphi$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  one violation for each pair  $(\mathbf{a}, \mathbf{b})$  of an underlying segment  $\mathbf{a}$  and a surface segment  $\mathbf{b}$  which are put in correspondence by  $\rho_{\mathbf{a},\mathbf{b}}$  despite the fact that they are assigned different values by the feature  $\varphi$  (McCarthy and Prince 1995). To illustrate,  $\text{IDENT}_{[\text{nasal}]}$  assigns two violations to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  in (74), because of the two corresponding pairs  $(/n/, [t])$  and  $(/k/, [\eta])$ .

$$(74) \quad \begin{array}{l} \mathbf{a} = \text{a n t a } \eta \text{ k} \\ \quad \quad \quad | \quad \backslash \quad \backslash \quad \backslash \quad \backslash \\ \mathbf{b} = \text{a t a } \eta \text{ } \end{array}$$

Let's consider two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  together with their composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ . Does the identity faithfulness constraint  $\text{IDENT}_{[\text{nasal}]}$  satisfy the  $\text{FIC}_{\text{comp}}$  (67)?

We can reason exactly as in the two preceding subsections 3.3.1 and 3.3.2. We assume that the antecedent of the  $\text{FIC}_{\text{comp}}$  holds, namely that the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  does not violate  $\text{IDENT}_{[\text{nasal}]}$ , as in (75a).

$$(75) \quad \begin{array}{lll} \text{a. } \text{ID}_{[\text{nas}]}(\mathbf{b}, \mathbf{c}) = 0 & \text{b. } \text{ID}_{[\text{nas}]}(\mathbf{a}, \mathbf{c}) = 1 & \text{c. } \text{ID}_{[\text{nas}]}(\mathbf{a}, \mathbf{b}) = 1 \\ \\ \mathbf{b} = \text{g a} & \mathbf{a} = \eta \text{ a} & \mathbf{a} = \eta \text{ a} \\ \quad \quad | \quad | & \quad \quad | \quad | & \quad \quad | \quad | \\ \mathbf{c} = \text{g a} & \mathbf{c} = \text{g a} & \mathbf{b} = \text{g a} \end{array}$$

Also, we assume that the left-hand side of the inequality in the consequent of the  $\text{FIC}_{\text{comp}}$  is larger than zero, namely that the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  does violate  $\text{IDENT}_{[\text{nasal}]}$ , as in (75b). By definition of the composition correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ ,  $\rho_{\mathbf{a},\mathbf{b}}$  must put in correspondence the underlying nasal  $/\eta/$  of  $\mathbf{a}$  with the surface oral  $[g]$  of  $\mathbf{b}$ , as represented in (75c). This says in turn that the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which figures in the right-hand side of the  $\text{FIC}_{\text{comp}}$  inequality violates  $\text{IDENT}_{[\text{nasal}]}$  as well, so that the inequality holds.

Also for  $\text{IDENT}_{[\text{nasal}]}$ , as for  $\text{DEP}$ , the  $\text{FIC}_{\text{comp}}$  requires no underlying segment to be broken by the correspondence relation  $\rho_{\mathbf{b},\mathbf{c}}$ , as shown by the counterexample (76), analogous to (73).

$$(76) \quad \begin{array}{lll} \text{a. } \text{ID}_{[\text{nas}]}(\mathbf{b}, \mathbf{c}) = 0 & \text{b. } \text{ID}_{[\text{nas}]}(\mathbf{a}, \mathbf{c}) = 2 & \text{c. } \text{ID}_{[\text{nas}]}(\mathbf{a}, \mathbf{b}) = 1 \\ \\ \mathbf{b} = \text{g a} & \mathbf{a} = \eta \text{ a} & \mathbf{a} = \eta \text{ a} \\ \quad \quad / \quad | \quad | & \quad \quad / \quad | \quad | & \quad \quad | \quad | \\ \mathbf{c} = \text{g g a} & \mathbf{c} = \text{g g a} & \mathbf{b} = \text{g a} \end{array}$$

The antecedent of the  $\text{FIC}_{\text{comp}}$  holds: gemination preserves nasality in the candidate in (76a) (which could correspond for instance to the Japanese loan  $[\text{fu.r}\acute{o}g.\text{gu}]$  of English *frog*; Kubozono et al. 2008). But the inequality in the consequent of the  $\text{FIC}_{\text{comp}}$  fails: the composition candidate (76b) violates  $\text{IDENT}_{[\text{nasal}]}$  twice because of the gemination while the candidate (76c) violates it only once, so that the left-hand side of the inequality exceeds the right-hand side.

These considerations extend from  $\text{IDENT}_{[\text{nasal}]}$  to the identity faithfulness constraint  $\text{IDENT}_{\varphi}$  corresponding to any feature  $\varphi$ , independently of whether it is binary or multivalued, as long as it is total. The case of partial features is indeed more delicate. Assume that the identity faithfulness constraint  $\text{IDENT}_{\varphi}$  corresponding to a partial feature  $\varphi$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  one violation for each pair  $(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}}$  of corresponding segments such that the feature  $\varphi$  is defined for both segments and assigns them a different value. Thus,  $\text{IDENT}_{\varphi}$  is not violated when the feature  $\varphi$  is undefined for at least one of the two segments.<sup>9</sup> To illustrate, suppose that the feature  $[\text{strident}]$  is only defined for coronals (Hayes 2009). The corresponding constraint  $\text{IDENT}_{[\text{strident}]}$  does not satisfy the  $\text{FIC}_{\text{comp}}$  (67), as shown by the counterexample (77).

<sup>9</sup> Another option is to let  $\text{IDENT}_{\varphi}$  assign one violation also when the feature  $\varphi$  is undefined for one and only one of the two segments in the corresponding pair  $(\mathbf{a}, \mathbf{b})$ . This definition of  $\text{IDENT}_{\varphi}$  effectively treats  $\varphi$  as a total but multi-valued feature. Under this approach, feature partiality raises no additional complications for the  $\text{FIC}_{\text{comp}}$ .

$$(77) \quad \begin{array}{lll} \text{a. } \text{ID}_{[\text{str}]}(\mathbf{b}, \mathbf{c})=0 & \text{b. } \text{ID}_{[\text{str}]}(\mathbf{a}, \mathbf{c})=1 & \text{c. } \text{ID}_{[\text{str}]}(\mathbf{a}, \mathbf{b})=0 \\ & \mathbf{a} = \theta \mathbf{a} & \mathbf{a} = \theta \mathbf{a} \\ & \begin{array}{c} | \\ | \\ | \end{array} & \begin{array}{c} | \\ | \\ | \end{array} \\ \mathbf{b} = \mathbf{f} \mathbf{a} & & \mathbf{b} = \mathbf{f} \mathbf{a} \\ \begin{array}{c} | \\ | \end{array} & & \\ \mathbf{c} = \mathbf{s} \mathbf{a} & \mathbf{c} = \mathbf{s} \mathbf{a} & \end{array}$$

The antecedent of the  $\text{FIC}_{\text{comp}}$  holds: the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  in (77a) does not violate  $\text{IDENT}_{[\text{strident}]}$ , because the underlying /f/ of  $\mathbf{b}$  is not coronal and thus undefined for stridency. The right-hand side of the  $\text{FIC}_{\text{comp}}$  inequality is equal to 0: the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in (77c) does not violate  $\text{IDENT}_{[\text{strident}]}$ , because again [f] is undefined for stridency. The  $\text{FIC}_{\text{comp}}$  inequality fails because its left-hand side is equal to 1: the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$  in (77b) does violate  $\text{IDENT}_{[\text{strident}]}$ , because of the two corresponding coronal segments /θ/ and [s]. These considerations lead to proposition 4, which will be refined in section 3.4 and proven in appendix A.3.

**PROPOSITION 4 (provisional)** *The identity faithfulness constraint  $\text{IDENT}_{\varphi}$  relative to a phonological feature  $\varphi$  satisfies the  $\text{FIC}_{\text{comp}}$  (67) provided no correspondence relation in the candidate set breaks any underlying segment and furthermore the feature  $\varphi$  is total relative to the candidate set.*  $\square$

### 3.4. Establishing the $\text{FIC}_{\text{comp}}$ : restricted constraints

The phonological literature has made use of *restricted* variants of MAX, DEP, and IDENT which are only offended when the deleted, epenthetic, or mismatching segments belong to a *privileged* segment set. Privilege can be determined by segmental quality or position in the string. This section investigates how these restricted constraints fair with respect to the  $\text{FIC}_{\text{comp}}$ .

#### 3.4.1. $\text{MAX}_R$

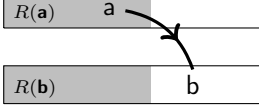
A *restriction*  $R$  pairs a string  $\mathbf{a}$  with a subset  $R(\mathbf{a})$  of its segments. A segment of the string  $\mathbf{a}$  *satisfies* the restriction provided it belongs to  $R(\mathbf{a})$ . The faithfulness constraint  $\text{MAX}_R$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each segment of the underlying string  $\mathbf{a}$  which satisfies the restriction  $R$  and is deleted. Deletion of underlying segments which do not satisfy the restriction is not penalized. To illustrate, consider the restriction  $R$  which pairs a string  $\mathbf{a}$  with the set  $R(\mathbf{a})$  of its consonants. The corresponding constraint  $\text{MAX}_R$  is the constraint MAX-C which militates against consonant deletion, but is not offended by vowel deletion (it thus assigns only one violation to the candidate (69), while unrestricted MAX assigns two violations).

While we have seen that unrestricted MAX satisfies the  $\text{FIC}_{\text{comp}}$ , its restricted counterpart  $\text{MAX}_R$  can fail at the  $\text{FIC}_{\text{comp}}$ , as shown by the counterexample (78) for  $\text{MAX}_R = \text{MAX-C}$ .

$$(78) \quad \begin{array}{lll} \text{a. } \text{MAX}_R(\mathbf{b}, \mathbf{c})=0 & \text{b. } \text{MAX}_R(\mathbf{a}, \mathbf{c})=1 & \text{c. } \text{MAX}_R(\mathbf{a}, \mathbf{b})=0 \\ & \mathbf{a} = \mathbf{e} \mathbf{s} & \mathbf{a} = \mathbf{e} \mathbf{s} \\ & \begin{array}{c} | \\ | \end{array} & \begin{array}{c} | \\ | \\ | \end{array} \\ \mathbf{b} = \mathbf{e} \mathbf{e} & & \mathbf{b} = \mathbf{e} \mathbf{e} \\ \begin{array}{c} | \\ | \end{array} & & \\ \mathbf{c} = \mathbf{e} & \mathbf{c} = \mathbf{e} & \end{array}$$

The antecedent of the  $\text{FIC}_{\text{comp}}$  holds: the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  in (78a) does not violate MAX-C, because the deleted segment is a vowel. The right-hand side of the  $\text{FIC}_{\text{comp}}$  inequality is equal to zero: the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in (78c) does not violate MAX-C, because it involves no deletion. The  $\text{FIC}_{\text{comp}}$  inequality thus fails, because its left-hand side is instead equal to 1: the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$  in (78b) does violate MAX-C, because it deletes a consonant.

In order for  $\text{MAX}_R$  to fail at the  $\text{FIC}_{\text{comp}}$  in (78), it is crucial that  $\rho_{\mathbf{a}, \mathbf{b}}$  establishes a correspondence between the consonant /s/ and the vowel [e], namely between a segment which satisfies the restriction  $R$  and a segment which does not satisfy it. Given a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  is said to *exit from* the restriction  $R$  if it puts some underlying segment  $\mathbf{a}$  which satisfies the restriction  $R$  in correspondence with some surface segment  $\mathbf{b}$  which does not satisfy it, as in (79). The top and bottom rectangles represent the sets of segments of  $\mathbf{a}$  and  $\mathbf{b}$ , with the subsets selected by the restriction  $R$  highlighted in gray.

$$(79) \quad (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, \quad \mathbf{a} \in R(\mathbf{a}), \quad \mathbf{b} \notin R(\mathbf{b})$$


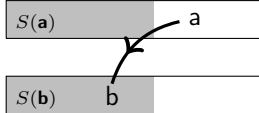
The following proposition ensures that  $\text{MAX}_R$  satisfies the  $\text{FIC}_{\text{comp}}$  provided no correspondence relation in the candidate set exits from the restriction  $R$ . To illustrate, the proposition guarantees that  $\text{MAX}_C$  satisfies the  $\text{FIC}_{\text{comp}}$  provided no underlying consonant is in correspondence with a surface vowel. The proposition will be further extended in subsection 3.4.3.

**PROPOSITION 2 (provisional)** *Assume the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose correspondence relation exits from the restriction  $R$ , so that condition (79) is impossible relative to the candidate set. The faithfulness constraint  $\text{MAX}_R$  then satisfies the  $\text{FIC}_{\text{comp}}$  (67).  $\square$*

A restriction is *trivial* provided it pairs every string with the totality of its segments. The case of unrestricted  $\text{MAX}$  discussed in subsection 3.3.1 follows as a special case of  $\text{MAX}_R$  with a trivial restriction  $R$ : no correspondence relation can exit from  $R$  in this case and (79) is thus contradictory.

### 3.4.2. $\text{DEP}^S$

The reasoning in subsection 3.4.1 extends straightforwardly from  $\text{MAX}$  to  $\text{DEP}$ . Given a restriction  $S$ , the corresponding faithfulness constraint  $\text{DEP}^S$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each segment of the surface string  $\mathbf{b}$  which satisfies the restriction  $S$  and is epenthetic.<sup>10</sup> To illustrate, consider the restriction  $S$  which pairs a string with the set of its vowels. The corresponding constraint  $\text{DEP}^S$  is the constraint  $\text{DEP-V}$  which militates against vowel epenthesis, but is not offended by consonant epenthesis. Given a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  is said to *enter into*  $S$  provided it puts some underlying segment  $\mathbf{a}$  which does not satisfy the restriction  $S$  in correspondence with some surface segment  $\mathbf{b}$  which does satisfy it, as in (80). Condition (80) is analogous to (79), only with the roles of underlying and surface segments switched.

$$(80) \quad (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, \quad \mathbf{a} \notin S(\mathbf{a}), \quad \mathbf{b} \in S(\mathbf{b})$$


In the case of  $\text{MAX}_R$ , the no-exiting assumption that (79) is impossible suffices to establish the  $\text{FIC}_{\text{comp}}$ . In the case of  $\text{DEP}^S$ , the no-entering assumption that (80) is impossible needs to be coupled with the no-breaking condition, as expected based on the discussion in subsection 3.3.2. The following proposition guarantees that it suffices to require the no-breaking condition among the segments which satisfy the restriction  $S$ , intuitively because  $\text{DEP}^S$  only cares about those segments. To illustrate, the proposition guarantees that  $\text{DEP-V}$  satisfies the  $\text{FIC}_{\text{comp}}$  provided no surface vowel is in correspondence with an underlying consonant and furthermore no underlying vowel is diphthongized.<sup>11</sup> The proposition will be further extended in subsection 3.4.4.

**PROPOSITION 3 (provisional)** *Assume that no underlying segment which satisfies the restriction  $S$  is broken into two surface segments which both satisfy the restriction  $S$ , in the sense that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  with two different corresponding pairs  $(\mathbf{a}, \mathbf{b}_1), (\mathbf{a}, \mathbf{b}_2) \in \rho_{\mathbf{a}, \mathbf{b}}$  for the same underlying segment  $\mathbf{a}$  such that  $\mathbf{a} \in S(\mathbf{a})$  and  $\mathbf{b}_1, \mathbf{b}_2 \in S(\mathbf{b})$ . Assume furthermore that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose correspondence relation enters into the restriction  $S$ , so that condition (80) is impossible relative to the candidate set. The faithfulness constraint  $\text{DEP}^S$  then satisfies the  $\text{FIC}_{\text{comp}}$  (67).  $\square$*

### 3.4.3. $\text{MAX}_R^S$

The *doubly restricted* constraint  $\text{MAX}_R^S$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each segment of the underlying string  $\mathbf{a}$  which satisfies the restriction  $R$  (namely, it belongs to  $R(\mathbf{a})$ ) and has no correspondent segment in the surface string  $\mathbf{b}$  which satisfies the restriction  $S$  (namely, it belongs to

<sup>10</sup> Throughout this section, a restriction on the underlying segments is denoted by  $R$  and appears as a subscript on the constraint's name (as in  $\text{MAX}_R$ ) while a restriction on the surface segments is denoted by  $S$  and appears as a superscript (as in  $\text{DEP}^S$ ).

<sup>11</sup> Assuming a breaking analysis of vowel diphthongization.

$S(\mathbf{b})$ —although it might have surface correspondents which do not satisfy the restriction  $S$ . To illustrate, consider the restriction  $R$  which pairs a string with the set of its consonants and the restriction  $S$  which pairs a string with the set of the segments in its initial syllable.  $\text{MAX}_R^S$  is Beckman’s (1999) constraint  $\text{MAX-C-}\sigma_1$  which mandates that every consonant has a correspondent in the initial syllable. The following proposition extends the analysis of the singly restricted  $\text{MAX}_R$  to the doubly restricted  $\text{MAX}_R^S$ . This proposition concludes the analysis of segmental  $\text{MAX}$  constraints. The proof is a straightforward verification, as shown in appendix A.1.

**Proposition 2.** *Assume that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  which satisfies condition (81), which is therefore impossible relative to the candidate set.*

$$(81) \quad (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, \quad \mathbf{a} \in R(\mathbf{a}), \quad \mathbf{b} \notin R(\mathbf{b}), \quad \mathbf{b} \in S(\mathbf{b})$$

The faithfulness constraint  $\text{MAX}_R^S$  then satisfies the  $FIC_{\text{comp}}$  (67).  $\square$

Condition (81) for  $\text{MAX}_R^S$  and condition (79) considered in subsection 3.4.1 for  $\text{MAX}_R$  differ only in that the former has the additional fourth clause  $\mathbf{b} \in S(\mathbf{b})$ . Because of this additional clause, the assumption that (81) is impossible required for  $\text{MAX}_R^S$  to satisfy the  $FIC_{\text{comp}}$  is weaker than the assumption that (79) is impossible required for  $\text{MAX}_R$ . To illustrate, the proposition says that the doubly restricted  $\text{MAX-C-}\sigma_1$  satisfies the  $FIC_{\text{comp}}$  provided no consonant is in correspondence with the vowel of the initial syllable—while the singly restricted  $\text{MAX-C}$  was shown in subsection 3.4.1 to require the stronger assumption that no consonant is in correspondence with any vowel.

The case of  $\text{MAX}_R$  follows as a special case of  $\text{MAX}_R^S$  with a trivial restriction  $S$ : the additional clause  $\mathbf{b} \in S(\mathbf{b})$  is trivially satisfied in this case and the two conditions (79) and (81) thus coincide. The constraint  $\text{MAX}^S$  always satisfies the  $FIC_{\text{comp}}$  because in this case the restriction  $R$  is trivially satisfied by every segment and thus the clause  $\mathbf{b} \notin R(\mathbf{b})$  is impossible. As another interesting special case, suppose that the two restrictions  $R$  and  $S$  coincide. Condition (81) is then contradictory, because it cannot be the case that  $\mathbf{b} \notin R(\mathbf{b})$  and  $\mathbf{b} \in S(\mathbf{b})$ . The constraint  $\text{MAX}_R^S$  thus satisfies the  $FIC_{\text{comp}}$  without additional assumptions. This observation will be used in subsections 3.5.1 to establish the  $FIC_{\text{comp}}$  for featural  $\text{MAX}$  constraints.

#### 3.4.4. $\text{DEP}_R^S$

The reasoning in subsection 3.4.3 extends straightforwardly from  $\text{MAX}_R^S$  to  $\text{DEP}_R^S$ . The doubly restricted constraint  $\text{DEP}_R^S$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each segment of the surface string  $\mathbf{b}$  which satisfies the restriction  $S$  (namely, it belongs to  $S(\mathbf{b})$ ) and has no correspondent segment in the underlying string  $\mathbf{a}$  which satisfies the restriction  $R$  (namely, it belongs to  $R(\mathbf{a})$ )—although it might have underlying correspondents which do not satisfy the restriction  $R$ . The following proposition extends the analysis of the singly restricted  $\text{DEP}^S$  to the doubly restricted  $\text{DEP}_R^S$  and thus concludes the analysis of segmental  $\text{DEP}$  constraints. The assumption that (82) is impossible is weaker than the assumption that (80) is impossible, because of the additional clause  $\mathbf{a} \in R(\mathbf{a})$ . The proof of the proposition is a straightforward verification, as shown in appendix A.2.

**Proposition 3.** *Assume that no underlying segment which satisfies the restriction  $S$  is broken into two surface segments which both satisfy the restriction  $S$ , in the sense that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  with two different corresponding pairs  $(\mathbf{a}, \mathbf{b}_1), (\mathbf{a}, \mathbf{b}_2) \in \rho_{\mathbf{a}, \mathbf{b}}$  for the same underlying segment  $\mathbf{a}$  such that  $\mathbf{a} \in S(\mathbf{a})$  and  $\mathbf{b}_1, \mathbf{b}_2 \in S(\mathbf{b})$ . Assume furthermore that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  which satisfies condition (82), which is therefore impossible relative to the candidate set.*

$$(82) \quad (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, \quad \mathbf{b} \in S(\mathbf{b}), \quad \mathbf{a} \notin S(\mathbf{a}), \quad \mathbf{a} \in R(\mathbf{a})$$

The faithfulness constraint  $\text{DEP}_R^S$  then satisfies the  $FIC_{\text{comp}}$  (67).  $\square$

#### 3.4.5. $\text{IDENT}_{\varphi, R}$

The faithfulness constraint  $\text{IDENT}_{\varphi, R}$  corresponding to a total feature  $\varphi$  and a restriction  $R$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each corresponding pair  $(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}$  of segments which differ for the value of the feature  $\varphi$  such that the underlying segment  $\mathbf{a}$  satisfies the restriction  $R$  (namely it belongs to  $R(\mathbf{a})$ ). To illustrate, consider the restriction  $R$  which pairs a string with the set of its nasal segments. The corresponding constraint  $\text{IDENT}_{[\text{nasal}], R}$  is the constraint  $\text{IDENTI} \rightarrow \text{O} [+ \text{nasal}]$  of Pater (1999), which

punishes de-nasalization (i.e., an underlying nasal segment with an oral surface correspondent), but not nasalization. Proposition 4/A guarantees that  $\text{IDENT}_{\varphi,R}$  satisfies the  $\text{FIC}_{\text{comp}}$  provided the candidate set makes (83) impossible (and furthermore satisfies the usual no-breaking assumption).

**Proposition 4 (A).** *Assume that no correspondence relation in the candidate set breaks any underlying segment. Consider a feature  $\varphi$  which is total relative to the candidate set. Assume furthermore that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which satisfies condition (83), which is therefore impossible relative to the candidate set.*

$$(83) \quad (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}}, \quad \mathbf{a} \in R(\mathbf{a}), \quad \mathbf{b} \notin R(\mathbf{b}), \quad \varphi(\mathbf{a}) = \varphi(\mathbf{b})$$

The faithfulness constraint  $\text{IDENT}_{\varphi,R}$  then satisfies the  $\text{FIC}_{\text{comp}}$  (67).  $\square$

Condition (83) coincides with condition (79) used above in the analysis of  $\text{MAX}_R$ , apart from the additional clause  $\varphi(\mathbf{a}) = \varphi(\mathbf{b})$  that the two segments  $\mathbf{a}$  and  $\mathbf{b}$  are assigned the same value by the feature  $\varphi$ . The assumption that (83) is impossible thus means that the correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$  cannot exit from  $R$  without changing the value of the feature  $\varphi$ . The assumption that (83) is impossible is thus weaker than the assumption that (79) is impossible, which was needed above for  $\text{MAX}_R$ . To illustrate, consider again the feature  $\varphi = [\text{nasal}]$  and the restriction  $R$  which pairs a string with the set of its nasals. Condition (83) is contradictory in this case, because the three clauses  $\mathbf{a} \in R(\mathbf{a})$ ,  $\mathbf{b} \notin R(\mathbf{b})$ , and  $\varphi(\mathbf{a}) = \varphi(\mathbf{b})$  cannot hold simultaneously. Pater's constraint  $\text{IDENTI} \rightarrow \text{O}[+\text{nasal}] = \text{IDENT}_{[\text{nasal}],R}$  thus satisfies the  $\text{FIC}_{\text{comp}}$  (provided there is no breaking).

### 3.4.6. $\text{IDENT}_{\varphi}^S$

The faithfulness constraint  $\text{IDENT}_{\varphi}^S$  is defined analogously, the only difference being that the restriction is applied to surface rather than underlying segments. To illustrate, consider the restriction  $S$  which pairs a string with the set of segments which belong to its initial syllable. The corresponding constraint  $\text{IDENT}_{[\text{high}]}^S$  is the constraint  $\text{IDENT}_{[\text{high}]}^{\sigma_1}$  of Beckman (1997, 1999), which is violated by a surface vowel in the initial syllable in correspondence with an underlying vowel which differs with respect to the feature [high]. As another example, consider the restriction  $S$  which pairs a string with the set of its nasal segments. The corresponding constraint  $\text{IDENT}_{[\text{nasal}]}^S$  is the constraint  $\text{IDENTO} \rightarrow \text{I}[+\text{nasal}]$  of Pater (1999), which punishes nasalization (i.e., an underlying oral segment with a nasal surface correspondent), but not de-nasalization. Proposition 4/B guarantees that  $\text{IDENT}_{\varphi}^S$  satisfies the  $\text{FIC}_{\text{comp}}$  provided condition (84) is impossible. This assumption means that the correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$  cannot exit from  $R$  without changing the value of the feature  $\varphi$ . The only difference between propositions 4/A and 4/B is that the no-breaking assumption in the latter proposition is restricted to the segments which satisfy the restriction. The proof of both propositions is a straightforward verification, as shown in appendix A.3.

**PROPOSITION 4 (B)** *Assume that no underlying segment which satisfies the restriction  $S$  can be broken into two surface segments which both satisfy the restriction  $S$ , in the sense that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  with two different candidate pairs  $(\mathbf{a}, \mathbf{b}_1), (\mathbf{a}, \mathbf{b}_2) \in \rho_{\mathbf{a},\mathbf{b}}$  for the same underlying segment  $\mathbf{a}$  such that  $\mathbf{a} \in S(\mathbf{a})$  and  $\mathbf{b}_1, \mathbf{b}_2 \in S(\mathbf{b})$ . Consider a feature  $\varphi$  which is total relative to the candidate set. Assume furthermore that the candidate set contains no candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which satisfies condition (84), which is therefore impossible relative to the candidate set.*

$$(84) \quad (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}}, \quad \mathbf{b} \in S(\mathbf{b}), \quad \mathbf{a} \notin S(\mathbf{a}), \quad \varphi(\mathbf{a}) = \varphi(\mathbf{b})$$

The faithfulness constraint  $\text{IDENT}_{\varphi}^S$  then satisfies the  $\text{FIC}_{\text{comp}}$  (67).  $\square$

### 3.4.7. $\text{IDENT}_{\varphi,R}^S$

For completeness, let's also consider the faithfulness constraint  $\text{IDENT}_{\varphi,R}^S$  corresponding to a total feature  $\varphi$  and two restrictions  $R, S$ , which assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  one violation for each pair  $(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}}$  of an underlying segment  $\mathbf{a}$  which satisfies the restriction  $R$  (namely, it belongs to  $R(\mathbf{a})$ ) and a surface segment  $\mathbf{b}$  which satisfies the restriction  $S$  (namely, it belongs to  $S(\mathbf{b})$ ) such that  $\varphi(\mathbf{a}) \neq \varphi(\mathbf{b})$ . To illustrate, consider the constraint  $\text{*REPLACE}(\text{coronal}, \text{labial})$  proposed in Boersma (1998): it is violated by an underlying coronal consonant with a labial surface correspondent. It can be reinterpreted as the constraint  $\text{IDENT}_{[\text{place}],R}^S$  corresponding to a tri-valued feature [place] where the restrictions  $R$  and  $S$  pair a string with the set of its coronal segments and the set of its labial segments, respectively. As

another example, consider the \*MAP constraint in (85), proposed by White (2013) and Hayes and White (to appear) building on Zuraw (2007, 2013): it is violated by an underlying voiceless stop which is in correspondence with a surface voiced fricative. This constraint can be reinterpreted as the constraint  $\text{IDENT}_{[\text{voice}],R}^S$  or  $\text{IDENT}_{[\text{cont}],R}^S$  where the restrictions  $R$  and  $S$  pair a string with the set of its voiceless stops and the set of its voiced fricatives, respectively.<sup>12</sup>

$$(85) \quad *_{\text{MAP}} \left( \left[ \begin{array}{c} -\text{voice} \\ -\text{cont} \end{array} \right], \left[ \begin{array}{c} +\text{voice} \\ +\text{cont} \end{array} \right] \right)$$

No simple conditions on the correspondence relations and the restrictions  $R, S$  seem to suffice to ensure that the doubly restricted constraint  $\text{IDENT}_{\varphi,R}^S$  satisfies the  $\text{FIC}_{\text{comp}}$ . In particular, it does not suffice to simply assume that the two conditions (83) and (84) for  $\text{IDENT}_{\varphi}^S$  and  $\text{IDENT}_{\varphi,R}$  are both impossible. Here is a counterexample. Consider a feature  $\varphi$  which is partial and binary. Consider the corresponding feature  $\widehat{\varphi}$  which is total and ternary, in the sense that  $\widehat{\varphi}$  coincides with  $\varphi$  for any segment that  $\varphi$  is defined for, while  $\widehat{\varphi}$  assigns the dummy value “0” to the segments that  $\varphi$  is undefined for. Consider the identity faithfulness constraint  $\text{IDENT}_{\varphi}$  relative to the partial feature  $\varphi$ , which only penalizes an underlying and a corresponding surface segments when the feature is defined for both and assigns them a different value ( $\text{IDENT}_{\varphi}$  does not assign a violation when the feature is defined for exactly one of the two corresponding segments). This constraint  $\text{IDENT}_{\varphi}$  is identical to the doubly restricted identity faithfulness constraint  $\text{IDENT}_{\widehat{\varphi},R}^S$  relative to the total three-valued feature  $\widehat{\varphi}$  and the restrictions  $R = S$  which pair a string with the set of its segments for which the feature  $\varphi$  is defined (namely the set of segments to which the corresponding total feature  $\widehat{\varphi}$  assigns values  $\pm$  and not the dummy value). Since the identity constraint  $\text{IDENT}_{\varphi}$  corresponding to the partial feature  $\varphi$  has been shown not to satisfy the  $\text{FIC}_{\text{comp}}$  in subsection 3.3.3, the doubly restricted constraint  $\text{IDENT}_{\widehat{\varphi},R}^S$  cannot satisfy the  $\text{FIC}_{\text{comp}}$  either. And yet conditions (83) and (84) are both contradictory, because the restrictions  $R, S$  are defined in terms of the values of the feature  $\widehat{\varphi}$ .

### 3.5. Establishing the $\text{FIC}_{\text{comp}}$ : other constraints

This section completes the analysis of the  $\text{FIC}_{\text{comp}}$  within Correspondence Theory, by looking at a variety of other faithfulness constraints which naturally arise within that framework. For simplicity, only the unrestricted versions of these constraints are considered.

#### 3.5.1. $\text{MAX}_{[+\varphi]}$ , $\text{DEP}_{[+\varphi]}$

Let “+” be a designated value of a feature  $\varphi$ , either partial or total, either binary or multi-valued. The faithfulness constraint  $\text{MAX}_{[+\varphi]}$  assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  one violation for each segment of the underlying string  $\mathbf{a}$  which has the designated value + for the feature  $\varphi$  but has no correspondent in the the surface string  $\mathbf{b}$  which shares the value + for the feature  $\varphi$  (Casali 1997, 1998; Walker 1999; Lombardi 2001).<sup>13</sup> To illustrate,  $\text{MAX}_{[+\text{voice}]}$  assigns two violations to the candidate (86), because both /b/ and /d/ lose their voicing (through devoicing and deletion, respectively).

$$(86) \quad \begin{array}{c} \mathbf{a} = \mathbf{b a d} \\ \quad \quad \quad \parallel \\ \mathbf{b} = \mathbf{p a} \end{array}$$

The featural constraint  $\text{MAX}_{[+\varphi]}$  coincides with the doubly restricted segmental constraint  $\text{MAX}_R^S$  when the two restrictions  $R$  and  $S$  both pair up a string with the set of its segments which have the value + for the feature  $\varphi$ . Since in particular  $R = S$ , condition (81) of proposition 2 is impossible, because its last two clauses  $\mathbf{b} \notin R(\mathbf{b})$  and  $\mathbf{b} \in S(\mathbf{b})$  are contradictory. The following result thus follows as a special case of proposition 2.

<sup>12</sup> Zuraw (2013) actually assumes that \*MAP applies to corresponding output segments. Output-output correspondence falls outside the scope of this work.

<sup>13</sup> Throughout this document, I assume correspondence relations to be defined among segments. Correspondence relations among feature values are then defined *indirectly*: two feature values are in correspondence provided their segmental carriers are in correspondence. The investigation of idempotency in an auto-segmental framework where correspondence relations are defined *directly* among feature values is left for future research.

**Proposition 5.** Let “+” be a designated value of a feature  $\varphi$  (either binary or multivalued, either partial or total). The faithfulness constraint  $\text{MAX}_{[+\varphi]}$  satisfies the  $\text{FIC}_{\text{comp}}$  (67) under no additional assumptions.  $\square$

Analogous considerations hold for the constraint  $\text{DEP}_{[+\varphi]}$ , which assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each segment of the surface string  $\mathbf{b}$  which has the designated value + for the feature  $\varphi$  but has no correspondent in the underlying string  $\mathbf{a}$  which shares the value +.

**Proposition 6.** Let “+” be a designated value of a feature  $\varphi$  (either binary or multivalued, either partial or total). Assume that no underlying segment with value + is broken into two surface segments which both have value +. The faithfulness constraint  $\text{DEP}_{[+\varphi]}$  then satisfies the  $\text{FIC}_{\text{comp}}$  (67).  $\square$

The featural constraints  $\text{MAX}_{[+\varphi]}/\text{DEP}_{[+\varphi]}$  differ subtly from the restricted segmental constraints  $\text{MAX}_R/\text{DEP}^S$  where  $R$  and  $S$  both pair a string with the set of its segments which have the value “+” for the feature  $\varphi$ . In fact,  $\text{MAX}_R/\text{DEP}^S$  are violated by an underlying/surface segment which has value “+” and is deleted/epenthesized, while  $\text{MAX}_{[+\varphi]}/\text{DEP}_{[+\varphi]}$  are violated by an underlying/surface segment which has the value “+” and is deleted/epenthesized or put in correspondence with segments with a different value for feature  $\varphi$ . This subtle difference is computationally substantial:  $\text{MAX}_{[+\varphi]}/\text{DEP}_{[+\varphi]}$  satisfy the  $\text{FIC}_{\text{comp}}$  under no additional assumptions; but  $\text{MAX}_R/\text{DEP}^S$  require the additional no-entering and no-exiting assumptions that (79) and (80) are impossible, as seen in subsections 3.4.1-3.4.2. Formally, this difference is due to the fact that  $\text{MAX}_{[+\varphi]}/\text{DEP}_{[+\varphi]}$  coincide not with  $\text{MAX}_R/\text{DEP}^S$  but with  $\text{MAX}_R^S/\text{DEP}_R^S$  with  $S = R$ .

### 3.5.2. UNIFORMITY and INTEGRITY

The faithfulness constraint UNIFORMITY assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each surface *coalescence*, namely for each segment of the surface string  $\mathbf{b}$  which has two or more correspondents in the underlying string  $\mathbf{a}$  according to  $\rho_{\mathbf{a}, \mathbf{b}}$  (McCarthy and Prince 1995). To illustrate, UNIFORMITY assigns two violations to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in (87), because of its two surface coalescences [b] and [f]. The constraint thus defined is *coarse* (Wheeler 2005): it does not distinguish between a coalescence of just two segments (such as [f] below) and a coalescence of more than two segments (such as [b]). This distinction can be captured through the following alternative *gradient* definition: the faithfulness constraint  $\text{UNIFORMITY}^{\text{grad}}$  assigns  $k$  violations for each coalescence of  $k \geq 2$  underlying segments.

$$(87) \quad \begin{array}{c} \mathbf{a} = \text{skwism} \\ \quad \searrow \quad \downarrow \quad \downarrow \quad \downarrow \\ \mathbf{b} = \quad \text{bif} \end{array}$$

While DEP punishes surface segments which have *no* underlying correspondents, UNIFORMITY punishes surface segments which have *too many*. The analysis of DEP in subsection 3.3.2 extends to UNIFORMITY, yielding the following proposition 7, whose simple verification is omitted for brevity.

**Proposition 7.** The faithfulness constraints UNIFORMITY and  $\text{UNIFORMITY}^{\text{grad}}$  satisfy the  $\text{FIC}_{\text{comp}}$  (67) provided no correspondence relation in the candidate set breaks any underlying segment.  $\square$

Analogous considerations hold for the faithfulness constraint INTEGRITY, which assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each *broken* underlying segment, namely for each segment of the underlying string  $\mathbf{a}$  which has two or more correspondents in the surface string  $\mathbf{b}$  according to  $\rho_{\mathbf{a}, \mathbf{b}}$  (see McCarthy and Prince 1995 and Staroverov 2014 for discussion). The corresponding gradient constraint  $\text{INTEGRITY}^{\text{grad}}$  assigns instead  $k$  violations for each underlying segment which is broken into  $k \geq 2$  surface segments. While MAX punishes underlying segments which have *no* surface correspondents, INTEGRITY punishes underlying segments which have *too many*. The analysis of MAX in subsection 3.3.1 extends to INTEGRITY, yielding the following proposition 8.

**Proposition 8.** The faithfulness constraints INTEGRITY and  $\text{INTEGRITY}^{\text{grad}}$  satisfy the  $\text{FIC}_{\text{comp}}$  (67) under no additional assumptions.  $\square$

### 3.5.3. CONTIGUITY

The faithfulness constraint I-CONTIGUITY assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  one violation for each *skipped* underlying segment, namely for each segment of the underlying string  $\mathbf{a}$  which has no correspondents in the surface string  $\mathbf{b}$  according to  $\rho_{\mathbf{a}, \mathbf{b}}$  and furthermore is flanked both on the left and on the right







$(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  in (90a) does not violate the conjoined constraint, because /e/ and [i] only differ for the feature [high]. The right-hand side of the  $\text{FIC}_{\text{comp}}$  inequality is small, namely equal to zero: the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in (90c) does not violate the conjoined constraint, because /a/ and [e] only differ for the feature [low]. The  $\text{FIC}_{\text{comp}}$  inequality fails because its left-hand side is large, namely equal to 1: the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}}, \rho_{\mathbf{b}, \mathbf{c}})$  in (90b) violates the conjoined constraint, because /a/ and [i] differ for both features [low] and [high].

$$(90) \quad \begin{array}{lll} \text{a. } \text{ID}_{[\text{low}]} \wedge \text{ID}_{[\text{high}]}(\mathbf{b}, \mathbf{c}) = 0 & \text{b. } \text{ID}_{[\text{low}]} \wedge \text{ID}_{[\text{high}]}(\mathbf{a}, \mathbf{c}) = 1 & \text{c. } \text{ID}_{[\text{low}]} \wedge \text{ID}_{[\text{high}]}(\mathbf{a}, \mathbf{b}) = 0 \\ \\ \begin{array}{c} \mathbf{b} = \text{e} \\ | \\ \mathbf{c} = \text{i} \end{array} & \begin{array}{c} \mathbf{a} = \text{a} \\ | \\ \mathbf{c} = \text{i} \end{array} & \begin{array}{c} \mathbf{a} = \text{a} \\ | \\ \mathbf{b} = \text{e} \end{array} \end{array}$$

The case of constraint disjunction is different. For concreteness, consider the disjunction  $\text{IDENT}_{\varphi} \vee \text{IDENT}_{\psi}$  of two identity faithfulness constraints  $\text{IDENT}_{\varphi}$  and  $\text{IDENT}_{\psi}$  corresponding to two (total) features  $\varphi$  and  $\psi$ . This constraint assigns one violation for each pair of corresponding segments which differ for either features  $\varphi$  or  $\psi$  (possibly both). Proposition 11 ensures that it satisfies the  $\text{FIC}_{\text{comp}}$ . The proof is a straightforward verification, which is omitted for brevity. This conclusion easily extends to the disjunction of other (disjoinable) faithfulness constraints: conditions on the  $\text{FIC}_{\text{comp}}$ -compliance of the constraint disjunction follow by combining conditions on the  $\text{FIC}_{\text{comp}}$ -compliance of the faithfulness constraints being combined in the disjunction.<sup>18</sup>

**Proposition 11.** *Assume that the features  $\varphi$  and  $\psi$  are total and that correspondence relations are not allowed to break any underlying segment. The disjunctive faithfulness constraint  $\text{IDENT}_{\varphi} \vee \text{IDENT}_{\psi}$  then satisfies the  $\text{FIC}_{\text{comp}}$  (67).  $\square$*

The difference between constraint conjunction and disjunction with respect to the  $\text{FIC}_{\text{comp}}$  can be appreciated as follows. Suppose that the antecedent of the  $\text{FIC}_{\text{comp}}$  holds for the disjunction  $\text{IDENT}_{[\text{low}]} \vee \text{IDENT}_{[\text{high}]}$ . This means that the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  does not violate it. This entails in turn that the candidate violates neither  $\text{IDENT}_{[\text{low}]}$  nor  $\text{IDENT}_{[\text{high}]}$ . The  $\text{FIC}_{\text{comp}}$  for the disjunction thus follows from the  $\text{FIC}_{\text{comp}}$  previously established for the individual disjuncts. The case of conjunction is different: even if the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  does not violate the conjunction  $\text{IDENT}_{[\text{low}]} \wedge \text{IDENT}_{[\text{high}]}$  as required by the antecedent of the  $\text{FIC}_{\text{comp}}$ , it could nonetheless violate one of the two conjuncts  $\text{IDENT}_{[\text{low}]}$  or  $\text{IDENT}_{[\text{high}]}$ . The fact that the conjuncts satisfy the  $\text{FIC}_{\text{comp}}$  thus provides no guarantees that their conjunction satisfies it as well.

### 3.6. Conclusions

A grammar is idempotent provided it faithfully maps to itself any phonotactically licit phonological form. Equivalently, a grammar fails at idempotency provided it displays at least one chain shift. Within constraint-based phonology, the typology of grammars is defined through a constraint set and a candidate set. Formal grammatical conditions such as idempotency must therefore be derivable from assumptions on the constraint set and assumptions on the candidate set, which exclude potentially dangerous constraints and candidates. This chapter has pursued this idea within the (classical) OT implementation of constraint-based phonology. Building on Tesar's (2013) theory of output-drivenness, the theory of idempotency has been developed in this chapter through two steps. First, proposition 1 has distilled the  $\text{FIC}_{\text{comp}}$  as a general condition on the faithfulness constraints which suffices to ensure idempotency. Second, propositions 2-11 have established the  $\text{FIC}_{\text{comp}}$  for a number of faithfulness constraints which naturally arise within McCarthy and Prince's (1995) Correspondence Theory, under various assumptions on the correspondence relations in the candidate set. The overall picture obtained by combining these propositions is summarized in the following result, which summarizes this chapter.

<sup>18</sup> The situation is rather different for the disjunction of a faithfulness and a markedness constraint. Such a disjunction can yield a constraint which is neither a faithfulness nor a markedness constraint (see Wolf 2007 for broader discussion), contrary to what required by the *classical* implementation of OT defined in subsection 3.2.1.

**Proposition 12.** Consider a candidate set which consists of triplets  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  of an underlying segment string  $\mathbf{a}$  and a surface segment string  $\mathbf{b}$  together with a correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$  between the segments of  $\mathbf{a}$  and those of  $\mathbf{b}$ . Assume that this candidate set satisfies the reflexivity axiom (51) and the transitivity axiom (66), repeated below in (91) and (92).

- (91) If the candidate set contains a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  with a surface form  $\mathbf{b}$ , it also contains the corresponding identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ , where  $\mathbb{I}_{\mathbf{b},\mathbf{b}}$  is the identity correspondence relation among the segments of  $\mathbf{b}$ .
- (92) If the candidate set contains two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which share  $\mathbf{b}$  as the surface and underlying form, it also contains the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  whose correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  is the composition of  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$ .

Assume that no correspondence relation breaks any underlying segment, namely puts it in correspondence with two or more surface segments. Assume furthermore that the constraint set only contains faithfulness constraints drawn from the following list:

- (93) a.  $\text{IDENT}_\varphi$  (when the feature  $\varphi$  is total, not necessarily binary);  
 b. segmental  $\text{MAX}$  and  $\text{DEP}$ ;  
 c. featural  $\text{MAX}_{[\pm\varphi]}$  and  $\text{DEP}_{[\pm\varphi]}$  (for any feature  $\varphi$ );  
 d.  $\text{UNIFORMITY}$ ,  $\text{LINEARITY}$  and  $\text{ADJACENCY}$ .

The OT grammar corresponding to any ranking of this constraint set is idempotent. This conclusion extends to the case where the constraint set contains restricted variants of these constraints (such as  $\text{MAX}_R$ ,  $\text{DEP}^S$ ,  $\text{IDENT}_{\varphi,R}$ , or  $\text{IDENT}_\varphi^S$ ), as long as no correspondence relation crosses the restrictions, namely puts in correspondence a segment which satisfies the restrictions with a segment which does not satisfy them.  $\square$

## A. Proofs

Throughout this appendix, I consider three strings  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$ , whose generic segments are denoted by  $a$ ,  $b$ , and  $c$ . I use statements such as “for every/some segment  $\mathbf{a}$ ” as a shorthand for “for every/some segment  $\mathbf{a}$  of the string  $\mathbf{a}$ ”, thus leaving the domain of the quantifiers implicit.

### A.1. Proof of proposition 2

A segment  $\mathbf{a}$  violates the constraint  $\text{MAX}_R^S$  relative to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  provided  $\mathbf{a}$  belongs to  $R(\mathbf{a})$  and it has no  $\rho_{\mathbf{a},\mathbf{b}}$ -correspondent in the surface string  $\mathbf{b}$  which belongs to  $S(\mathbf{b})$ .  $\text{MAX}_R^S$  assigns one violation for each underlying segment which violates it. This appendix proves proposition 2 repeated below, which establishes the  $\text{FIC}_{\text{comp}}$  for  $\text{MAX}_R^S$ .

**PROPOSITION 2.** The faithfulness constraint  $\text{MAX}_R^S$  satisfies the  $\text{FIC}_{\text{comp}}$

- (94) **If:**  $\text{MAX}_R^S(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$   
**Then:**  $\text{MAX}_R^S(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq \text{MAX}_R^S(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$

for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  such that the former candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  has no underlying segment  $\mathbf{a}$  and no surface segment  $\mathbf{b}$  such that:

- (95)  $(a, b) \in \rho_{\mathbf{a},\mathbf{b}}, \quad a \in R(\mathbf{a}), \quad b \notin R(\mathbf{b}), \quad b \in S(\mathbf{b})$   $\square$

**PROOF.** Assume that the antecedent of the implication (94) holds, namely that the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  does not violate  $\text{MAX}_R^S$ . The following chain of implications (96) then holds for any segment  $\mathbf{a}$  of the string  $\mathbf{a}$ . In step (96a), I have used the definition of  $\text{MAX}_R^S$ . In step (96b), I have used the definition of the composition correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ . In step (96c), I have used the antecedent of the implication (94), which guarantees that (\*) entails (\*\*). In fact, suppose by contradiction that (\*) is true but (\*\*) false. This means that  $\mathbf{b} \in R(\mathbf{b})$  and furthermore  $\mathbf{b}$  has no surface correspondent  $\mathbf{c}$  which belongs to  $S(\mathbf{c})$ . In other words, the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  incurs at least one violation of  $\text{MAX}_R^S$ , contradicting the antecedent of the implication (94). In step (96d), I have replaced (\*\*) with (\*\*\*) because of the assumption that (95) is impossible. In step (96e), I have used again the definition of  $\text{MAX}_R^S$ .

$$\begin{aligned}
(96) \quad & \mathbf{a} \text{ violates } \text{MAX}_R^S \text{ relative to } (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \\
& \xLeftrightarrow{(a)} \mathbf{a} \in R(\mathbf{a}) \text{ and } \forall \mathbf{c} \left[ (\mathbf{a}, \mathbf{c}) \in \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}} \rightarrow \mathbf{c} \notin S(\mathbf{c}) \right] \\
& \xLeftrightarrow{(b)} \mathbf{a} \in R(\mathbf{a}) \text{ and } \forall \mathbf{b} \left[ (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \rightarrow \underbrace{\forall \mathbf{c} \left[ (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b},\mathbf{c}} \rightarrow \mathbf{c} \notin S(\mathbf{c}) \right]}_{(*)} \right] \\
& \xLeftrightarrow{(c)} \mathbf{a} \in R(\mathbf{a}) \text{ and } \forall \mathbf{b} \left[ (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \rightarrow \underbrace{\mathbf{b} \notin R(\mathbf{b})}_{(**)} \right] \\
& \xLeftrightarrow{(d)} \mathbf{a} \in R(\mathbf{a}) \text{ and } \forall \mathbf{b} \left[ (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \rightarrow \underbrace{\mathbf{b} \notin S(\mathbf{b})}_{(***)} \right] \\
& \xLeftrightarrow{(e)} \mathbf{a} \text{ violates } \text{MAX}_R^S \text{ relative to } (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})
\end{aligned}$$

The chain of implications (96) says that, if an underlying segment  $\mathbf{a}$  violates  $\text{MAX}_R^S$  relative to the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ , then  $\mathbf{a}$  also violates  $\text{MAX}_R^S$  relative to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . This conclusion establishes the inequality in the consequent of the  $\text{FIC}_{\text{comp}}$  (94).  $\square$

By (96), if  $\mathbf{a}$  violates  $\text{MAX}_R^S$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ , then  $\mathbf{a}$  also violates  $\text{MAX}_R^S$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . Suppose that the restriction  $S$  is trivial, namely  $S$  pairs any string  $\mathbf{a}$  with the complete set  $S(\mathbf{a})$  of its segments. In this case, the reverse implication trivially holds as well: if  $\mathbf{a}$  violates  $\text{MAX}_R$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ , then  $\mathbf{a}$  has no  $\rho_{\mathbf{a},\mathbf{b}}$ -correspondents in  $\mathbf{b}$  and therefore  $\mathbf{a}$  cannot have any  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ -correspondents in  $\mathbf{c}$  neither, thus violating  $\text{MAX}_R$  also relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ . In conclusion, the inequality in the consequent of the  $\text{FIC}_{\text{comp}}$  (94) holds as an identity in the case where  $S$  is the trivial restriction.

## A.2. Proof of proposition 3

A segment  $\mathbf{b}$  violates the constraint  $\text{DEP}_R^S$  relative to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  provided  $\mathbf{b}$  belongs to  $S(\mathbf{b})$  and has no  $\rho_{\mathbf{a},\mathbf{b}}$ -correspondent in the underlying string  $\mathbf{a}$  which belongs to  $R(\mathbf{a})$ .  $\text{DEP}_R^S$  assigns one violation for each surface segment which violates it. This appendix proves proposition 3 repeated below, which establishes the  $\text{FIC}_{\text{comp}}$  for  $\text{DEP}_R^S$ .

**PROPOSITION 3.** *The faithfulness constraint  $\text{DEP}_R^S$  satisfies the  $\text{FIC}_{\text{comp}}$*

$$\begin{aligned}
(97) \quad \text{If:} \quad & \text{DEP}_R^S(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0 \\
\text{Then:} \quad & \text{DEP}_R^S(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq \text{DEP}_R^S(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})
\end{aligned}$$

for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  such that the latter candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  has no underlying segment  $\mathbf{b}$  and no surface segment  $\mathbf{c}$  such that

$$(98) \quad (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b},\mathbf{c}}, \quad \mathbf{c} \in S(\mathbf{c}), \quad \mathbf{b} \notin S(\mathbf{b}), \quad \mathbf{b} \in R(\mathbf{b})$$

and furthermore it enforces no breaking among the segments which satisfy the restriction  $S$ , namely there exist no underlying segment  $\mathbf{b}$  and no surface segments  $\mathbf{c}_1, \mathbf{c}_2$  such that

$$(99) \quad \mathbf{b} \in S(\mathbf{b}), \quad \mathbf{c}_1, \mathbf{c}_2 \in S(\mathbf{c}), \quad (\mathbf{b}, \mathbf{c}_1) \in \rho_{\mathbf{b},\mathbf{c}}, \quad (\mathbf{b}, \mathbf{c}_2) \in \rho_{\mathbf{b},\mathbf{c}}, \quad \mathbf{c}_1 \neq \mathbf{c}_2 \quad \square$$

PROOF. Assume that the antecedent of the implication (97) holds, namely that the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  does not violate  $\text{DEP}_R^S$ . The following chain of implications (100) then holds for any segment  $\mathbf{c}$  of the string  $\mathbf{c}$ . In step (100a), I have used the definition of  $\text{DEP}_R^S$ . In step (100b), I have used the definition of the composition correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ . In step (100c), I have used the antecedent of the implication (97), which guarantees that the surface segment  $\mathbf{c} \in S(\mathbf{c})$  admits a correspondent underlying segment  $\mathbf{b} \in R(\mathbf{b})$  according to  $\rho_{\mathbf{b},\mathbf{c}}$ . In step (100d), I have replaced  $(*)$  with  $(**)$  because of the assumption that (98) is impossible. In step (100e), I have used again the definition of  $\text{DEP}_R^S$ . Note that the no-breaking assumption that (99) is impossible has not been used in deriving the chain of implications (100). Note that the chain of implications (100) makes no use of the no-breaking assumption (99).

$$\begin{aligned}
(100) \quad & \text{c violates } \text{DEP}_R^S \text{ relative to } (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \\
& \stackrel{(a)}{\iff} \text{c} \in S(\mathbf{c}) \text{ and } \forall \mathbf{a} \left[ (\mathbf{a}, \mathbf{c}) \in \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}} \rightarrow \mathbf{a} \notin R(\mathbf{a}) \right] \\
& \stackrel{(b)}{\iff} \text{c} \in S(\mathbf{c}) \text{ and } \forall \mathbf{b} \left[ (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b},\mathbf{c}} \rightarrow \forall \mathbf{a} \left[ (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \rightarrow \mathbf{a} \notin R(\mathbf{a}) \right] \right] \\
& \stackrel{(c)}{\implies} \text{c} \in S(\mathbf{c}) \text{ and } \exists \mathbf{b} \left[ (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b},\mathbf{c}} \text{ and } \underbrace{\mathbf{b} \in R(\mathbf{b})}_{(*)} \text{ and } \forall \mathbf{a} \left[ (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \rightarrow \mathbf{a} \notin R(\mathbf{a}) \right] \right] \\
& \stackrel{(d)}{\implies} \text{c} \in S(\mathbf{c}) \text{ and } \exists \mathbf{b} \left[ (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b},\mathbf{c}} \text{ and } \underbrace{\mathbf{b} \in S(\mathbf{b})}_{(**)} \text{ and } \forall \mathbf{a} \left[ (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \rightarrow \mathbf{a} \notin R(\mathbf{a}) \right] \right] \\
& \stackrel{(e)}{\iff} \text{c} \in S(\mathbf{c}) \text{ and } \exists \mathbf{b} \left[ (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b},\mathbf{c}} \text{ and } \mathbf{b} \text{ violates } \text{DEP}_R^S \text{ relative to } (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \right]
\end{aligned}$$

By (100), for each segment  $\mathbf{c}$  which violates  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ , there exists a segment  $\mathbf{b}$  which violates  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  such that  $(\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b},\mathbf{c}}$ . This is not sufficient to secure the inequality in the consequent of (97). In fact, it could still be the case that two *different* segments  $\mathbf{c}_1$  and  $\mathbf{c}_2$  which violate  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  both correspond to the *same* segment  $\mathbf{b}$  which violates  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . In this case,  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  could have more epenthetic segments than  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and the inequality in the consequent of (97) would fail.<sup>19</sup> In order to secure the inequality in the consequent of (97), I need to make sure that the mapping from the segments  $\mathbf{c}_{1,2}, \dots$  which violate  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  to the segments  $\mathbf{b}_1, \mathbf{b}_2, \dots$  which violate  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  defined by (100) is *injective*: if two violating segments  $\mathbf{c}_1, \mathbf{c}_2$  are different, the two corresponding violating segments  $\mathbf{b}_1, \mathbf{b}_2$  are different as well.

The no-breaking assumption (99) serves precisely this purpose. Indeed, consider two different segments  $\mathbf{c}_1$  and  $\mathbf{c}_2$  which both violate the constraint  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  and thus both belong to  $S(\mathbf{c})$ . By (100), there exist segments  $\mathbf{b}_1, \mathbf{b}_2$  such that  $(\mathbf{b}_1, \mathbf{c}_1), (\mathbf{b}_2, \mathbf{c}_2) \in \rho_{\mathbf{b},\mathbf{c}}$  and furthermore  $\mathbf{b}_1, \mathbf{b}_2$  violate  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and thus belong to  $S(\mathbf{b})$ . If it were  $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}$ , it would be  $(\mathbf{b}, \mathbf{c}_1) \in \rho_{\mathbf{b},\mathbf{c}}$  and  $(\mathbf{b}, \mathbf{c}_2) \in \rho_{\mathbf{b},\mathbf{c}}$ , contradicting the no-breaking assumption (99). In conclusion, (100) establishes an *injective* mapping from the segments  $\mathbf{c}_1, \mathbf{c}_2, \dots$  which violate  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  to the segments  $\mathbf{b}_1, \mathbf{b}_2, \dots$  which violate  $\text{DEP}_R^S$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . This conclusion establishes the inequality in the consequent of the  $\text{FIC}_{\text{comp}}$  (97).  $\square$

The two chains of implications (96) and (100) in the proofs of the two propositions 2 and 3 for  $\text{MAX}_R^S$  and  $\text{DEP}_R^S$  are completely analogous. Yet, the no-crossing assumptions (95) and (98) target different correspondence relations: the former targets the correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$ ; the latter targets the correspondence relation  $\rho_{\mathbf{b},\mathbf{c}}$ .

### A.3. Proof of proposition 4

A pair  $(\mathbf{a}, \mathbf{b})$  of an underlying segment  $\mathbf{a}$  and a surface segment  $\mathbf{b}$  violate the faithfulness constraint  $\text{IDENT}_{\varphi,R}$  relative to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  provided the following three conditions hold: the two segments  $\mathbf{a}$  and  $\mathbf{b}$  are in correspondence:  $(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}}$ ; the two segments  $\mathbf{a}$  and  $\mathbf{b}$  differ with respect to feature  $\varphi$ :  $\varphi(\mathbf{a}) \neq \varphi(\mathbf{b})$ ; the underlying segment  $\mathbf{a}$  belongs to  $R(\mathbf{a})$  and thus satisfies the restriction  $R$ .  $\text{IDENT}_{\varphi,R}$  assigns one violation for each underlying/surface segment pair  $(\mathbf{a}, \mathbf{b})$  which violates it. This appendix proves proposition 4/A repeated below, which establishes the  $\text{FIC}_{\text{comp}}$  for  $\text{IDENT}_{\varphi,R}$ . The proof of proposition 4/B for  $\text{IDENT}_{\varphi}^S$  is analogous.

**PROPOSITION 4 (A)** *Consider a total feature  $\varphi$ . The identity faithfulness constraint  $\text{IDENT}_{\varphi,R}$  satisfies the  $\text{FIC}_{\text{comp}}$*

$$\begin{aligned}
(101) \quad & \text{If: } \text{IDENT}_{\varphi,R}(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0 \\
& \text{Then: } \text{IDENT}_{\varphi,R}(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq \text{IDENT}_{\varphi,R}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})
\end{aligned}$$

for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  such that the former candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  has no underlying segment  $\mathbf{a}$  and no surface segment  $\mathbf{b}$  such that

<sup>19</sup> That is precisely what happens in the counterexample (73): the two segments  $\mathbf{c}_1 = [\text{e}]$  and  $\mathbf{c}_2 = [\text{i}]$  are both epenthetic relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ ; they both correspond to the same segment  $\mathbf{b} = [\text{ə}]$  which is indeed epenthetic relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . The candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  thus has more epenthetic segments than the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and the inequality in the consequent of (97) thus fails.

$$(102) \quad (\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, \quad \mathbf{a} \in R(\mathbf{a}), \quad \mathbf{b} \notin R(\mathbf{b}), \quad \varphi(\mathbf{a}) = \varphi(\mathbf{b})$$

and furthermore either the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  does not break any underlying segment into two or more surface segments, namely there exist no underlying segment  $\mathbf{a}$  and no surface segments  $\mathbf{b}_1, \mathbf{b}_2$  such that

$$(103) \quad (\mathbf{a}, \mathbf{b}_1) \in \rho_{\mathbf{a}, \mathbf{b}}, \quad (\mathbf{a}, \mathbf{b}_2) \in \rho_{\mathbf{a}, \mathbf{b}}, \quad \mathbf{b}_1 \neq \mathbf{b}_2 \quad \square$$

PROOF. Assume that the antecedent of the implication (101) holds, namely that the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  does not violate  $\text{IDENT}_{\varphi, R}$ . The following chain of implications (104) then holds for any segment  $\mathbf{a}$  of the string  $\mathbf{a}$  and any segment  $\mathbf{c}$  of the string  $\mathbf{c}$ . In step (104a), I have used the definition of the constraint  $\text{IDENT}_{\varphi, R}$ . In step (104b), I have used the definition of the composition correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}}$ . In step (104c), I have added the conjunct (\*). This is licit because the antecedent of (101) guarantees that the the pair of segments  $(\mathbf{b}, \mathbf{c})$  does not violate the constraint  $\text{IDENT}_{\varphi, R}$  relative to  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ . Since they are in correspondence through  $\rho_{\mathbf{b}, \mathbf{c}}$ , this means that either  $\varphi(\mathbf{b}) \notin R(\mathbf{b})$  or else  $\varphi(\mathbf{b}) = \varphi(\mathbf{c})$ . In step (104d), I have replaced (\*) with the equivalent (\*\*). In fact, if it is the first condition  $\varphi(\mathbf{b}) = \varphi(\mathbf{c})$  of the disjunction (\*) which holds, then it must be  $\varphi(\mathbf{a}) \neq \varphi(\mathbf{b})$ , because  $\varphi(\mathbf{a}) \neq \varphi(\mathbf{c})$ . If it is instead the second condition  $\mathbf{b} \notin R(\mathbf{b})$  of the disjunction (\*) which holds, then it must be  $\varphi(\mathbf{a}) \neq \varphi(\mathbf{b})$ , because (102) is impossible. Finally, in step (104e), I have used again the definition of  $\text{IDENT}_{\varphi, R}$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . Note that the chain of implications (104) does not make use of the no-breaking assumption (103).

$$\begin{aligned}
(104) \quad & (\mathbf{a}, \mathbf{c}) \text{ violates } \text{IDENT}_{\varphi, R} \text{ relative to } (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}}) \\
& \stackrel{(a)}{\iff} \varphi(\mathbf{a}) \in R(\mathbf{a}), \varphi(\mathbf{a}) \neq \varphi(\mathbf{c}), (\mathbf{a}, \mathbf{c}) \in \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}} \\
& \stackrel{(b)}{\iff} \varphi(\mathbf{a}) \in R(\mathbf{a}), \varphi(\mathbf{a}) \neq \varphi(\mathbf{c}), \exists \mathbf{b} [(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b}, \mathbf{c}}] \\
& \stackrel{(c)}{\iff} \varphi(\mathbf{a}) \in R(\mathbf{a}), \varphi(\mathbf{a}) \neq \varphi(\mathbf{c}), \exists \mathbf{b} [(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b}, \mathbf{c}}, \underbrace{[\varphi(\mathbf{b}) = \varphi(\mathbf{c}) \text{ or } \mathbf{b} \notin R(\mathbf{b})]}_{(*)}] \\
& \stackrel{(d)}{\iff} \varphi(\mathbf{a}) \in R(\mathbf{a}), \varphi(\mathbf{a}) \neq \varphi(\mathbf{c}), \exists \mathbf{b} [(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a}, \mathbf{b}}, (\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b}, \mathbf{c}}, \underbrace{\varphi(\mathbf{a}) \neq \varphi(\mathbf{b})}_{(**)}] \\
& \stackrel{(e)}{\iff} \exists \mathbf{b} [(\mathbf{b}, \mathbf{c}) \in \rho_{\mathbf{b}, \mathbf{c}} \text{ and } (\mathbf{a}, \mathbf{b}) \text{ violates } \text{IDENT}_{\varphi, R} \text{ relative to } (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})]
\end{aligned}$$

Consider two different pairs  $(\mathbf{a}_1, \mathbf{c}_1)$  and  $(\mathbf{a}_2, \mathbf{c}_2)$  which both violate the faithfulness constraint  $\text{IDENT}_{\varphi, R}$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$ . The chain of implications (104) guarantees that there exist  $\mathbf{b}_1, \mathbf{b}_2$  such that the two pairs  $(\mathbf{a}_1, \mathbf{b}_1)$  and  $(\mathbf{a}_2, \mathbf{b}_2)$  violate the faithfulness constraint  $\text{IDENT}_{\varphi, R}$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and furthermore  $(\mathbf{b}_1, \mathbf{c}_1) \in \rho_{\mathbf{b}, \mathbf{c}}$  and  $(\mathbf{b}_2, \mathbf{c}_2) \in \rho_{\mathbf{b}, \mathbf{c}}$ . If  $\mathbf{a}_1 \neq \mathbf{a}_2$ , also the two pairs  $(\mathbf{a}_1, \mathbf{b}_1)$  and  $(\mathbf{a}_2, \mathbf{b}_2)$  are different. Thus assume that  $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}$ , whereby  $\mathbf{c}_1 \neq \mathbf{c}_2$ . If it were  $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}$ , then the latter two conditions would say that  $(\mathbf{b}, \mathbf{c}_1) \in \rho_{\mathbf{b}, \mathbf{c}}$  and  $(\mathbf{b}, \mathbf{c}_2) \in \rho_{\mathbf{b}, \mathbf{c}}$ , contradicting the hypothesis that the correspondence relation  $\rho_{\mathbf{b}, \mathbf{c}}$  cannot break any underlying segment  $\mathbf{b}$ . In conclusion, (104) defines an injective mapping from the pairs  $(\mathbf{a}, \mathbf{c})$  which violate the constraint  $\text{IDENT}_{\varphi, R}$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  to the pairs  $(\mathbf{a}, \mathbf{b})$  which violate it relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . This conclusion establishes the inequality in the consequent of the implication (101).  $\square$



## Idempotency and the triangle inequality

A phonological grammar is idempotent provided it maps any phonotactically licit form faithfully to itself. Chapter 3 has shown that idempotency in OT follows from a condition on the faithfulness constraints, referred to as the FIC. Since phonology is by-and-large idempotent (chain shifts are attested but marginal), faithfulness constraint must by-and-large satisfy the FIC, as I have indeed verified in chapter 3 for a variety of constraints which naturally arise within the Correspondence Theory of faithfulness. That is apparently surprising, since the FIC looks *prima facie* like a rather technical formal condition: why should the faithfulness constraints used in phonological theory satisfy it? This chapter answers this question by showing that the FIC actually admits an intuitive interpretation: it simply requires the faithfulness constraints to measure the phonological distance between underlying and surface representation in a sensible way, namely in compliance with the metrical *triangle inequality*.

This chapter is organized as follows. Section 4.1 introduces this crucial metric inequality for faithfulness constraints. Section 4.2 offers a digression on Harmonic Grammar, where the connection between idempotency and the triangle inequality is more readily visible. Section 4.3 probes deeper into the formal underpinning of the theory of faithfulness, developing a thorough formalization of McCarthy’s (2003) conjecture that faithfulness constraints are *categorical*. Finally, section 4.4 shows that the FIC for categorical faithfulness constraints is indeed equivalent to the requirement that they measure phonological distances in compliance with the triangle inequality.

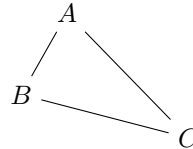
### 4.1. The metric nature of the faithfulness constraints

This section formalizes the intuition that faithfulness constraints measure the phonological distance between underlying and surface forms in compliance with a core axiom of the notion of distance, namely the *triangle inequality*.

#### 4.1.1. Intuition: the metric triangle inequality

A *distance* (or a *metric*) maps two points  $A$  and  $B$  to a non-negative value  $dist(A, B)$ . In order to capture the intuitive notion of distance, this mapping needs to satisfy some core axioms (Rudin 1953, ch. 2). One of these axioms is the *triangle inequality* (105): the distance between two points  $A$  and  $C$  is never larger than the sum of the distance between  $A$  and  $B$  plus the distance between  $B$  and  $C$ , no matter how the intermediate point  $B$  is chosen. In other words, no side of a triangle can be longer than the sum of the other two sides.

$$(105) \quad dist(A, C) \leq dist(A, B) + dist(B, C)$$



Motivated by the intuition that a faithfulness constraint measures the phonological distance between underlying and corresponding surface forms, I extend the metric triangle inequality to faithfulness constraints as condition (106) for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  which share the string  $\mathbf{b}$  as the underlying and the surface form, respectively.<sup>1</sup>

$$(106) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$$

In order to enforce this condition, we need an explicit assumption on the correspondence relation  $\rho_{\mathbf{a}, \mathbf{c}}$  which appears on the left-hand side of (106).

<sup>1</sup> The triangle inequality is applied here to each faithfulness constraint individually, rather than to some aggregation (e.g., a weighted sum) of their number of violations.



### 4.1.2. Formalization: the $\text{FTI}_{\text{comp}}$

No faithfulness constraint satisfies condition (106) for *every* correspondence relation  $\rho_{\mathbf{a},\mathbf{c}}$ , because some choices of  $\rho_{\mathbf{a},\mathbf{c}}$  make the left-hand side too large (for instance when  $\rho_{\mathbf{a},\mathbf{c}}$  is empty and  $F$  is MAX or DEP). Conversely, every faithfulness constraint satisfies condition (106) for *some* correspondence relation  $\rho_{\mathbf{a},\mathbf{c}}$ , because some choices of  $\rho_{\mathbf{a},\mathbf{c}}$  make the left-hand side equal to zero (for instance when  $\rho_{\mathbf{a},\mathbf{c}}$  is total and  $F$  is MAX or DEP). Some link is needed between the correspondence relations  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$  on the right-hand side of (106) and the correspondence relation  $\rho_{\mathbf{a},\mathbf{c}}$  on the left-hand side. A natural assumption is that  $\rho_{\mathbf{a},\mathbf{c}}$  is the composition  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  of the correspondence relations  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$ , whose existence is guaranteed by the transitivity axiom (66). In conclusion, a faithfulness constraint  $F$  is said to satisfy the *faithfulness triangle inequality* ( $\text{FTI}_{\text{comp}}$ ) provided condition (107) holds for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  and their composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ . The subscript “comp” in the acronym  $\text{FTI}_{\text{comp}}$  makes it explicit that the candidate on the left-hand side of the inequality (107) is the composition candidate.

$$(107) \quad F(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}_{\rho_{\mathbf{a},\mathbf{c}}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$$

Subsection 4.4.6 will show that a wide range of faithfulness constraints within Correspondence Theory satisfy the  $\text{FTI}_{\text{comp}}$ .

### 4.1.3. Faithfulness constraints and other metrical axioms

Besides the triangle inequality, a function *dist* from pair of points to non-negative numbers is required to satisfy two additional axioms in order to satisfy the abstract definition of distance. The first axiom is *symmetry*, which requires the distance between two points to be insensitive to their order:  $\text{dist}(A, B) = \text{dist}(B, A)$ . The second axiom is the *identity of the indiscernibles*, which requires two points to coincide if and only if their distance is null:  $A = B$  iff  $\text{dist}(A, B) = 0$ . Together with the triangle inequality, these axioms ensure the non-negativity of a distance:  $\text{dist}(A, B) \geq 0$ . Symmetry holds for certain faithfulness constraints (most notably, those of the IDENT type), but fails for core faithfulness constraints such as MAX and DEP (see Magri 2015 for the computational implications of symmetric faithfulness constraints). Half of the identity of the indiscernibles is enforced by the definition (56) of the faithfulness constraints provided in subsection 3.2.1: if the underlying and surface forms coincide (and the correspondence relation is the identity), their faithfulness violations are equal to zero. But the other half of the identity of the indiscernibles fails, as faithfulness constraints can obviously be satisfied by less than perfect string identity. These considerations motivate the focus on the remaining crucial metric axiom, namely the triangle inequality.

### 4.1.4. Summary

This section has introduced the  $\text{FTI}_{\text{comp}}$  (107), which formalizes the intuition that faithfulness constraints measure the phonological distance between underlying and surface forms in compliance with a crucial axiom of the abstract notion of distance, namely the triangle inequality. A number of questions now arises: which faithfulness constraints satisfy this metrical faithfulness condition? what are its implications for phonological theory. The rest of this chapter will address these questions.

## 4.2. Idempotency in Harmonic Grammar and the triangle inequality

Chapter 3 has introduced the structural condition of idempotency and has investigated it for (classical) OT. In particular, it has established the  $\text{FIC}_{\text{comp}}$  as a sufficient condition for idempotency in OT. The goal of this chapter is to provide an intuitive interpretation of the  $\text{FIC}_{\text{comp}}$ , by showing that it is related to the  $\text{FTI}_{\text{comp}}$ , namely to the requirement that faithfulness constraints measure phonological distance in a sensible way, namely in compliance with the triangle inequality. Towards establishing this connection, this section makes a digression on the related framework of *Harmonic Grammar* (HG; Legendre et al. 1990b,a; Smolensky and Legendre 2006), which allows for a more pristine view of the relationship between idempotency and the faithfulness triangle inequality.

### 4.2.1. Harmonic Grammar

As OT, also HG presupposes an underlying candidate set together with a constraint set which extracts the relevant phonological properties of the candidates. A *constraint weighting*  $\theta$  assigns a numerical

weight  $\theta_C \geq 0$  to each constraint  $C$ . The weighting  $\theta$  *prefers* a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  to another candidate  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}})$  provided the weighted sum of the constraint violations of the former candidate is smaller than that of the latter, as in (108). The HG grammar  $G_\theta$  corresponding to a constraint weighting  $\theta$  maps an underlying form  $\mathbf{a}$  to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which is preferred by the weighting  $\theta$  to all other candidates  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  which share that underlying form  $\mathbf{a}$ .

$$(108) \quad \sum_C \theta_C C(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) < \sum_C \theta_C C(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}})$$

Constraints are always interpreted as expressing penalties, never rewards. Hence, constraint weights need to be enforced to satisfy the non-negativity condition  $\theta_C \geq 0$  in order for HG to avoid undesired typological predictions, whereby less marked structures are mapped to more marked ones (Legendre et al. 2006; Keller 2000). This assumption that constraint weights are non-negative plays a crucial role in the rest of this section (in particular in the proof of Proposition 13 below provided in Appendix A.1).

#### 4.2.2. A condition on the candidate set: the one-to-one assumption

Which conditions on the candidate and the constraint set ensure that the HG grammars corresponding to any weighting is idempotent? Let me start with candidate conditions. The counterexamples in subsection 3.3.2 and 3.3.3 have shown that OT idempotency fails when correspondence relations can break an underlying segment into multiple surface segments. Let me recall the counterexample. The grammar (109) fails at idempotency:  $\mathbf{b} = [\mathbf{g}]$  is phonotactically licit, as it is the surface realization of  $\mathbf{a} = /[\eta]/$ , which IDENT<sub>[nasal]</sub> prevents from geminating; yet  $\mathbf{b} = /[\mathbf{g}]/$  cannot surface faithfully, because IDENT<sub>[nasal]</sub> fails at protecting it against gemination.

(109)

	/η/	IDENT <sub>[nasal]</sub>	STRESS
☞ $\mathbf{a} = \eta$		*	*
$\mathbf{b} = \mathbf{g}$			
$\mathbf{a} = \eta$		**!	
$\mathbf{c} = \mathbf{g}\mathbf{g}$			

	/g/	IDENT <sub>[nasal]</sub>	STRESS
$\mathbf{b} = \mathbf{g}$			*!
$\mathbf{b} = \mathbf{g}$			
☞ $\mathbf{b} = \mathbf{g}$			
$\mathbf{c} = \mathbf{g}\mathbf{g}$			

The *no-breaking assumption* is thus necessary for idempotency in OT. The same counterexample of course shows that the no-breaking assumption is needed for idempotency in HG as well.

The situation is different for coalescence of two underlying segments into a single surface segment. While coalescence does not threaten OT idempotency, it does hinder HG idempotency, as shown by the following counterexample. Before the suffix *-i-*, Finnish long low vowels shorten (*/aa/* → *[a]*), short low vowels raise (*/a/* → *[o]*), short round vowels surface faithfully (*/o/* → *[o]*) (Łubowicz 2011).

(110)

		<i>sing nom</i>	<i>plural essive</i>	
a.	<i>/aa/</i> → <i>[a]</i> :	<b>maa</b>	<b>ma-i-na</b>	‘earth’
		<b>vapaa</b>	<b>vapa-i-na</b>	‘free’
b.	<i>/a/</i> → <i>[o]</i> :	<b>kissa</b>	<b>kisso-i-na</b>	‘cat’
		<b>vapa</b>	<b>vapo-i-na</b>	‘fishing rod’
c.	<i>/o/</i> → <i>[o]</i> :	<b>talo</b>	<b>talo-i-na</b>	‘house’
		<b>pelko</b>	<b>pelko-i-na</b>	‘fear’

Let’s analyze vowel shortening as coalescence of two underlying vowels into a single surface vowel. The mapping */aa/* → *[a]* thus violates the faithfulness constraint  $F = \text{IDENT}_{[\text{low}]}$  twice, while the mapping */a/* → *[o]* violates it only once, as represented in (111). Assume that the constraint set also contains a markedness constraint  $M$  which punishes *[a]*. The HG grammar corresponding to the weights  $\theta_F = 2$  and  $\theta_M = 3$  yields the chain shift (110) and thus fails at idempotency: the short low vowel *[a]* is phonotactically licit, as it is the surface realization of the long low vowel */aa/* which IDENT<sub>[low]</sub> prevents from raising; yet, */a/* does not surface faithfully, because IDENT<sub>[low]</sub> fails to protect it.

(111)

	/aa/	$F = \text{IDENT}_{[\text{low}]}$	$M$
☞	<b>[a]</b>		*
	<b>[o]</b>	**	

	/a/	$F = \text{IDENT}_{[\text{low}]}$	$M$
	<b>[a]</b>		*
☞	<b>[o]</b>	*	

In conclusion, HG idempotency fails even in the simplest cases when correspondence relations display breaking or coalescence, thus effectively requiring the correspondence relations in the candidate set to be one-to-one.

#### 4.2.3. A condition on the faithfulness constraint set: the $\text{FIC}_{\text{comp}}^{\text{HG}}$

The following Proposition 13 provides the HG analogue of the OT Proposition 1. The proof is provided in Appendix A.1. Condition (112) will be referred to as the *HG faithfulness idempotency condition* ( $\text{FIC}_{\text{comp}}^{\text{HG}}$ ). No assumptions are made on the markedness constraints, on the nature of the faithfulness constraints (for instance, they are not required to be categorical), on the correspondence relations (for instance, they are not required to be one-to-one).

**Proposition 13.** *Assume that, for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , the candidate set also contains a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  such that the following condition (112) holds for any faithfulness constraint  $F$  in the constraint set.*

(112) For every choice of the constant  $\xi \geq 0$ :<sup>2</sup>

$$\text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi$$

$$\text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \xi$$

Then, the HG grammar corresponding to any weighting of the constraint set is idempotent, no matter what the markedness constraints look like.  $\square$   $\square$

In subsection 3.2.4, I have specialized the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  to the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ , by focusing on the composition candidate. Analogously, the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  (112) can be specialized to the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  (113), by choosing as the candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  the composition  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  of  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , whose existence is guaranteed by the transitivity axiom (66).

(113) For every choice of the constant  $\xi \geq 0$ :

$$\text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi$$

$$\text{Then: } F(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}_{\rho_{\mathbf{a},\mathbf{c}}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \xi$$

This sufficient condition for idempotency provided by the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  is *tight*: any faithfulness constraint which fails the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  can be shown to admit an elementary counterexample where HG idempotency fails.

#### 4.2.4. The $\text{FIC}_{\text{comp}}^{\text{HG}}$ is stronger than the $\text{FIC}_{\text{comp}}^{\text{OT}}$ in the general case

The sufficient condition for HG idempotency provided by the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  (113) entails the sufficient condition for OT idempotency provided by the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ , repeated in (114) for ease of comparison. In fact, the latter is a special case of the former corresponding to the choice  $\xi = 0$ .

(114) **If:**  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$

$$\text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$$

The reverse implication fails in the general case, showing that the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  is stronger than the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ . Here is a counterexample. The faithfulness constraint  $\text{IDENT}_{[\text{low}]}$  satisfies the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  under the no-breaking assumption, by Proposition ??b. Yet, the counterexample (111) shows that it fails at the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  (113) for the two corresponding candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  in (115a) and (115b), their composition  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  in (115c), and the constant  $\xi = 1.5$ .

(115) a.  $\text{IDT}_{[\text{low}]}(\mathbf{a}, \mathbf{b}) = 0$       b.  $\text{IDT}_{[\text{low}]}(\mathbf{b}, \mathbf{c}) = 1$       c.  $\text{IDT}_{[\text{low}]}(\mathbf{a}, \mathbf{c}) = 2$

$$\begin{array}{l} \mathbf{a} = \text{a a} \\ \mathbf{b} = \text{a} \end{array} \quad \begin{array}{l} \swarrow \\ \searrow \end{array}$$

$$\begin{array}{l} \mathbf{b} = \text{a} \\ \mathbf{c} = \text{o} \end{array} \quad \begin{array}{l} | \\ | \end{array}$$

$$\begin{array}{l} \mathbf{a} = \text{a a} \\ \mathbf{c} = \text{o} \end{array} \quad \begin{array}{l} \swarrow \\ \searrow \end{array}$$

The fact that the  $\text{FIC}_{\text{comp}}^{\text{HG}}$  is stronger than the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  in the general case is unsurprising, given that HG typologies are larger than OT typologies and thus harder to discipline to idempotency.

<sup>2</sup> The implication (112) trivially holds also for  $\xi < 0$ , because the antecedent is always false in that case, due to the non-negativity of constraint violations.

#### 4.2.5. The $FIC_{\text{comp}}^{\text{HG}}$ is equivalent to the $FTI_{\text{comp}}$ in the general case

Section 4.1 has introduced the  $FTI_{\text{comp}}$  repeated in (116) to formalize the condition that faithfulness constraints measure phonological distance in compliance with the triangle inequality. Suppose the antecedent of the  $FIC_{\text{comp}}^{\text{HG}}$  (113) holds. The term (\*) in (116) is thus smaller than  $\xi$  and (116) entails the consequent of the  $FIC_{\text{comp}}^{\text{HG}}$ . In other words, the  $FTI_{\text{comp}}$  entails the  $FIC_{\text{comp}}^{\text{HG}}$ .

$$(116) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \underbrace{F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})}_{(*)}$$

The reverse entailment holds as well, because the antecedent of the  $FIC_{\text{comp}}^{\text{HG}}$  (113) trivially holds with the position  $\xi = F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  and its consequent is equivalent to the  $FTI_{\text{comp}}$  (116) in this case. In conclusion, we have obtained the equivalence between the  $FIC_{\text{comp}}^{\text{HG}}$  and the  $FTI_{\text{comp}}$  stated by the following Proposition 14.

**Proposition 14.** *The  $FIC_{\text{comp}}^{\text{HG}}$  (113) and the  $FTI_{\text{comp}}$  repeated in (116) are equivalent: a faithfulness constraint satisfies the former relative to two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  if and only if it satisfies the latter.  $\square$*

Since the triangle inequality characterizes the intuitive notion of distance, this proposition provides an intuitive interpretation of the sufficient condition for HG idempotency provided by the  $FIC_{\text{comp}}^{\text{HG}}$  (113): the  $FIC_{\text{comp}}^{\text{HG}}$  simply requires the faithfulness constraints to measure the phonological distance between underlying and surface forms in a sensible way, namely in compliance with the triangle inequality.

#### 4.2.6. Summary

This section has extended the theory of idempotency from OT to HG. In the case of HG, the sufficient condition for idempotency provided by the  $FIC_{\text{comp}}^{\text{HG}}$  admits a straightforward metric interpretation in terms of the triangle inequality. I now would like to derive an analogous interpretation for the sufficient condition for idempotency in OT provided by the  $FIC_{\text{comp}}^{\text{OT}}$  derived in chapter 3. That turns out to require a closer look at the formal theory of faithfulness, developed in the next section.

### 4.3. Faithfulness categoricity, additivity and monotonicity

McCarthy (2003b) conjectures that phonological constraints are all categorical. Informally, this means that they assign at most one violation per *locus* of violation. McCarthy provides an explicit formalization of categoricity for markedness constraints (see his scheme (1) on p. 77). His treatment of faithfulness constraints is not as explicit: he discusses individual faithfulness constraints but does not provide a general scheme. This section fills the gap. The idea of the formalization is that a candidate can be sliced into “sub-candidates” by slicing either its correspondence relation, or its underlying string, or its surface string. A faithfulness constraint is additive provided the number of violations it assigns to a candidate is the sum of the number of violations it assigns to the sub-candidates. An additive faithfulness constraint is categorical provided it only takes values 0 or 1 when applied to the sub-candidates. Finally, additivity entails monotonicity, which says that the number of violations of bigger candidates (namely, those with more sub-candidates) ought to be larger.

#### 4.3.1. Additivity, categoricity, and monotonicity w.r.t. correspondence relations

Intuitively, the identity faithfulness constraint  $\text{IDENT}_{\varphi}$  corresponding to a (total) feature  $\varphi$  counts over pairs of corresponding segments (McCarthy and Prince 1995). This intuition can be formalized through the following two-step definition. First, the constraint is defined for a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  whose correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$  consists of a single segment pair  $(\mathbf{a}, \mathbf{b})$ , as in (117a). Second, the definition is extended to an arbitrary candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  by summing over pairs  $(\mathbf{a}, \mathbf{b})$  in the correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$ , as in (117b).

$$(117) \quad \begin{aligned} \text{a. } \text{IDENT}_{\varphi}(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}, \mathbf{b})\}) &= \begin{cases} 1 & \text{if } \mathbf{a} \text{ and } \mathbf{b} \text{ differ relative to feature } \varphi \\ 0 & \text{otherwise} \end{cases} \\ \text{b. } \text{IDENT}_{\varphi}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) &= \sum_{(\mathbf{a},\mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}}} \text{IDENT}_{\varphi}(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}, \mathbf{b})\}) \end{aligned}$$

Intuitively, the faithfulness constraints LINEARITY (McCarthy and Prince 1995) and I/O-ADJACENCY (Carpenter 2002) count over two pairs of corresponding segments. This intuition can be formalized through the following two-step definition. First, the constraints are defined for a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  whose correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$  consists of only two pairs of segments  $(\mathbf{a}_1, \mathbf{b}_1)$  and  $(\mathbf{a}_2, \mathbf{b}_2)$ , as in (118). I also consider the variant I/O-ADJACENCY<sup>grad</sup> in (119), which is sensitive to the number of intervening segments.

$$\begin{aligned}
(118) \quad & \text{LINEARITY}(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)\}) = \\
& = \begin{cases} 1 & \mathbf{a}_1 \text{ precedes } \mathbf{a}_2 \text{ in } \mathbf{a} \text{ but } \mathbf{b}_1 \text{ follows } \mathbf{b}_2 \text{ in } \mathbf{b} \\ 0 & \text{otherwise} \end{cases} \\
& \text{I-ADJACENCY}(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)\}) = \\
& = \begin{cases} 1 & \text{if } \mathbf{a}_1, \mathbf{a}_2 \text{ are adjacent in } \mathbf{a} \text{ but } \mathbf{b}_1, \mathbf{b}_2 \text{ are not adjacent in } \mathbf{b} \\ 0 & \text{otherwise} \end{cases} \\
& \text{O-ADJACENCY}(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)\}) = \\
& = \begin{cases} 1 & \text{if } \mathbf{b}_1, \mathbf{b}_2 \text{ are adjacent in } \mathbf{b} \text{ but } \mathbf{a}_1, \mathbf{a}_2 \text{ are not adjacent in } \mathbf{a} \\ 0 & \text{otherwise} \end{cases} \\
(119) \quad & \text{I-ADJACENCY}^{\text{grad}}(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)\}) = \\
& = \begin{cases} k & \text{if } \mathbf{a}_1, \mathbf{a}_2 \text{ are adjacent but } \mathbf{b}_1, \mathbf{b}_2 \text{ are separated by } k \text{ segments} \\ 0 & \text{otherwise} \end{cases} \\
& \text{O-ADJACENCY}^{\text{grad}}(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)\}) = \\
& = \begin{cases} k & \text{if } \mathbf{b}_1, \mathbf{b}_2 \text{ are adjacent but } \mathbf{a}_1, \mathbf{a}_2 \text{ are separated by } k \text{ segments} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Second, the constraints are defined for an arbitrary candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  by summing over any two pairs  $(\mathbf{a}_1, \mathbf{b}_1)$  and  $(\mathbf{a}_2, \mathbf{b}_2)$  in the relation  $\rho_{\mathbf{a},\mathbf{b}}$ , as in (120).

$$(120) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) = \sum_{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2) \in \rho_{\mathbf{a},\mathbf{b}}} F(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)\})$$

Generalizing from these examples, a faithfulness constraint  $F$  is called *additive relative to the correspondence relations* (C-additive) of order  $\ell$  provided the identity (121) holds for any candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . The sum is over subsets of cardinality  $\ell$  of the correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$  (not necessarily disjoint one from the other). In other words, the number of violations assigned by  $F$  to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  is the sum of the number of violations that  $F$  assigns to the “sub-candidates”  $(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), \dots, (\mathbf{a}_\ell, \mathbf{b}_\ell)\})$  obtained by considering all correspondence sub-relations  $\{(\mathbf{a}_1, \mathbf{b}_1), \dots, (\mathbf{a}_\ell, \mathbf{b}_\ell)\}$  of cardinality  $\ell$  of  $\rho_{\mathbf{a},\mathbf{b}}$ .

$$(121) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) = \sum_{(\mathbf{a}_1, \mathbf{b}_1), \dots, (\mathbf{a}_\ell, \mathbf{b}_\ell) \in \rho_{\mathbf{a},\mathbf{b}}} F(\mathbf{a}, \mathbf{b}, \{(\mathbf{a}_1, \mathbf{b}_1), \dots, (\mathbf{a}_\ell, \mathbf{b}_\ell)\})$$

$F$  is called *categorical relative to the correspondence relations* (C-categorical) of order  $\ell$  provided it is additive and furthermore can only take values 0 or 1 when it is applied to a candidate whose correspondence relation has cardinality  $\ell$ , so that the addenda on the right-hand side of (121) are all equal to 0 or 1. To illustrate, the constraint IDENT<sub>φ</sub> is C-categorical of order 1; the constraints I/O-ADJACENCY and LINEARITY are C-categorical of order 2; the constraints I/O-ADJACENCY<sup>grad</sup> are C-additive but not C-categorical.

Finally, a faithfulness constraint  $F$  is called *monotone relative to the correspondence relations* (C-monotone) provided it satisfies the implication (122): if two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{a}, \mathbf{b}, \widehat{\rho}_{\mathbf{a},\mathbf{b}})$  share both the underlying string  $\mathbf{a}$  and the surface string  $\mathbf{b}$  and only differ because the correspondence relation of the former is a subset of the correspondence relation of the latter,  $F$  assigns less violations to the former candidate than to the latter.

$$(122) \quad \begin{array}{ll} \mathbf{If:} & \rho_{\mathbf{a},\mathbf{b}} \subseteq \widehat{\rho}_{\mathbf{a},\mathbf{b}} \\ \mathbf{Then:} & F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \leq F(\mathbf{a}, \mathbf{b}, \widehat{\rho}_{\mathbf{a},\mathbf{b}}) \end{array}$$

Any C-additive faithfulness constraint is C-monotone, as a larger correspondence relation yields more addenda on the right-hand side of (121).

### 4.3.2. Additivity, categoricity, and monotonicity w.r.t. underlying strings

Intuitively, the faithfulness constraints MAX and INTEGRITY (McCarthy and Prince 1995), INTEGRITY<sup>grad</sup> (Wheeler 2005) and MAX<sub>[+ $\varphi$ ]</sub> (Casali 1997, 1998; Walker 1999; Lombardi 2001) count over underlying segments. This intuition can be formalized through the following two-step definition. First, these constraints are defined for a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose underlying string consists of a single segment  $\mathbf{a}$ , as in (123).

$$(123) \quad \begin{aligned} \text{MAX}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} 1 & \text{if } \mathbf{a} \text{ has no surface correspondent w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \\ 0 & \text{otherwise} \end{cases} \\ \text{INT}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} 1 & \text{if } \mathbf{a} \text{ has multiple correspondents w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \\ 0 & \text{otherwise} \end{cases} \\ \text{INT}^{\text{grad}}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} k & \text{if } \mathbf{a} \text{ has } k \geq 2 \text{ correspondents w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \\ 0 & \text{otherwise} \end{cases} \\ \text{MAX}_{[+\varphi]}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} 1 & \text{if } \mathbf{a} \text{ has value } + \text{ for feature } \varphi \text{ but } \mathbf{a} \text{ has no correspondent } \mathbf{b} \text{ w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \text{ such} \\ & \text{that } \mathbf{b} \text{ also has value } + \text{ for feature } \varphi \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Second, the constraints are defined for a generic candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  by summing over each segment  $\mathbf{a}$  of the underlying string  $\mathbf{a}$ , as in (124).<sup>3</sup> The correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b})}$  on the right-hand side of (124) is the restriction of the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  from the string  $\mathbf{a}$  to its segment  $\mathbf{a}$ .<sup>4</sup>

$$(124) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{a} \subseteq \mathbf{a}} F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b})})$$

Intuitively, the faithfulness constraint MAXLINEARITY (Heinz 2005) counts over subsequences of length 2 of the underlying string. This intuition can be formalized through the following two-step definition. First, the constraint is defined for a candidate  $(\mathbf{a}_1 \mathbf{a}_2, \mathbf{b}, \rho_{\mathbf{a}_1 \mathbf{a}_2, \mathbf{b}})$  whose underlying string  $\mathbf{a}_1 \mathbf{a}_2$  has length 2, as in (125a). Second, the constraint is defined for an arbitrary candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  by summing over all subsequences  $\mathbf{a}_1 \mathbf{a}_2$  of length 2 of the underlying string  $\mathbf{a}$ , as in (125b).

$$(125) \quad \begin{aligned} \text{a. MAXLIN}(\mathbf{a}_1 \mathbf{a}_2, \mathbf{b}, \rho_{\mathbf{a}_1 \mathbf{a}_2, \mathbf{b}}) &= \begin{cases} 1 & \text{if } \mathbf{a}_1, \mathbf{a}_2 \text{ have no correspondents } \mathbf{b}_1, \mathbf{b}_2 \text{ such that } \mathbf{b}_1 \text{ precedes } \mathbf{b}_2 \\ 0 & \text{otherwise} \end{cases} \\ \text{b. MAXLIN}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \sum_{\mathbf{a}_1 \mathbf{a}_2 \subseteq \mathbf{a}} \text{MAXLIN}(\mathbf{a}_1 \mathbf{a}_2, \mathbf{b}, \rho_{\mathbf{a}_1 \mathbf{a}_2, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \mathbf{a}_2, \mathbf{b})}) \end{aligned}$$

Generalizing from these examples, a faithfulness constraint  $F$  is called *additive relative to the underlying strings* (I-additive) of order  $\ell$  provided the identity (126) holds for any candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . The sum is over all (possibly overlapping) subsequences of  $\mathbf{a}$  of length  $\ell$ .<sup>5</sup>

$$(126) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{a}_1 \cdots \mathbf{a}_\ell \subseteq \mathbf{a}} F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}, \rho_{\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b})})$$

<sup>3</sup> For any two strings  $\mathbf{a}$  and  $\mathbf{b}$ , the notation  $\mathbf{a} \subseteq \mathbf{b}$  indicates that  $\mathbf{a}$  is a subsequence of  $\mathbf{b}$ :  $\mathbf{a}$  is obtained from  $\mathbf{b}$  by replacing some symbols of  $\mathbf{b}$  with the empty symbol.

<sup>4</sup> In other words,  $\rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b})}$  is the set of those pairs  $(\mathbf{a}', \mathbf{b}')$  in  $\rho_{\mathbf{a}, \mathbf{b}}$  such that  $\mathbf{a}' = \mathbf{a}$ . Once the underlying string  $\mathbf{a}$  is restricted to a single segment  $\mathbf{a}$ , the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  must necessarily be restricted to  $\rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b})}$ . In fact, the triplet  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  (with the singleton underlying segment  $\mathbf{a}$  and the original correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$ ) does not count as a candidate according to the assumption (49) that correspondence relations hold between the segments of the two strings in the candidate.

<sup>5</sup> A *subsequence* needs not consist of contiguous elements, contrary to a *substring*: *itk* is both a subsequence and a substring of *pitkol*, while *ptkl* is a subsequence but not a substring. It might be possible to define I-additivity in terms of a sum over *sub-strings* of contiguous segments, rather than over *sub-sequences* of possibly non-contiguous strings. For LINEARITY-type constraints, this modification would capture Heinz's (2005) proposal that only *immediate* precedence matters in the definition of the faithfulness constraints. Switching from sub-sequences to sub-strings would also have implications for ADJACENCY-type constraints. They have been defined in (118) and (120) as counting over corresponding pairs, whereby they qualify as C-categorical. An alternative definition of, say, I-ADJACENCY would be the following: it assigns to a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  a number of violations equal to the number of underlying adjacent pairs of segments which have no adjacent surface correspondents. If I-additivity is redefined in terms of sub-strings, then I-ADJACENCY qualifies as I-additive (it would not count as I-additive according to definition in terms of sub-sequences).

$F$  is called *categorical relative to the underlying strings* (I-categorical) of order  $\ell$  provided it is additive and furthermore it can only take values 0 or 1 when it is applied to a candidate whose underlying string has length  $\ell$ , so that the addenda on the right-hand side of (126) are all equal to 0 or 1. To illustrate, the constraints MAX, INTEGRITY, and MAX $_{[+\varphi]}$  are I-categorical of order 1; the constraint MAXLINEARITY is I-categorical of order 2; and the constraint INTEGRITY<sup>grad</sup> is I-additive but not I-categorical.

Finally, a faithfulness constraint  $F$  is called *monotone relative to the underlying strings* (I-monotone) provided it satisfies the implication (127): if two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\widehat{\mathbf{a}}, \mathbf{b}, \rho_{\widehat{\mathbf{a}}, \mathbf{b}})$  share both the surface string  $\mathbf{b}$  and the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  and only differ because the underlying string  $\mathbf{a}$  of the former is a subsequence of the underlying string  $\widehat{\mathbf{a}}$  of the latter,<sup>6</sup> then  $F$  assigns less violations to the former candidate than to the latter.

$$(127) \quad \begin{array}{ll} \text{If:} & \mathbf{a} \subseteq \widehat{\mathbf{a}} \\ \text{Then:} & F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) \leq F(\widehat{\mathbf{a}}, \mathbf{b}, \rho_{\widehat{\mathbf{a}}, \mathbf{b}}) \end{array}$$

Any I-additive faithfulness constraint is I-monotone, as a longer underlying string yields more addenda on the right-hand side of (126).

#### 4.3.3. A remark on the proper definition of I-additivity

Let me entertain an alternative definition of I-additivity through the alternative condition (128) instead of the actual condition (126) considered above. In both the actual condition (126) and the variant (128), the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  which appears on the left-hand side is replaced on the right-hand side with its restriction  $\rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \dots \mathbf{a}_\ell, \mathbf{b})}$  to the underlying subsequence  $\mathbf{a}_1 \dots \mathbf{a}_\ell$ . Yet, the actual condition (126) and the variant (128) differ because only the former also replaces the underlying string  $\mathbf{a}$  with the underlying subsequence  $\mathbf{a}_1 \dots \mathbf{a}_\ell$ . The variant (128) thus makes I-additivity completely analogous to condition (121) used to define C-additivity: in both cases, the right-hand side sums over candidates which share the underlying and surface strings of the original candidate, but have a restricted correspondence relation.

$$(128) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{a}_1 \dots \mathbf{a}_\ell \subseteq \mathbf{a}} F(\mathbf{a}_1 \dots \mathbf{a}_\ell, \mathbf{b}, \rho_{\mathbf{a}_1 \dots \mathbf{a}_\ell, \mathbf{b}})$$

Yet, this alternative condition (128) makes no sense. To see that, suppose that the original correspondence relation is empty:  $\rho_{\mathbf{a}, \mathbf{b}} = \emptyset$ . Its restriction to any underlying subsequence is thus empty as well:  $\rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \dots \mathbf{a}_\ell, \mathbf{b})} = \emptyset$ . Since the original and the restricted correspondence relations thus coincide, the original candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  which appears on the left-hand side of (128) and the candidates  $(\mathbf{a}_1 \dots \mathbf{a}_\ell, \mathbf{b}, \rho_{\mathbf{a}_1 \dots \mathbf{a}_\ell, \mathbf{b}})$  which appear on the right-hand side coincide and must therefore be assigned the same number of violations by the faithfulness constraint  $F$ . Condition (128) is thus contradictory, unless  $F$  never assigns any violations in the case of empty correspondence relations (which is obviously false for instance in the case of MAX).

#### 4.3.4. Additivity, categoricity, and monotonicity w.r.t. surface strings

Intuitively, the faithfulness constraints DEP, UNIFORMITY, UNIFORMITY<sup>grad</sup>, and DEP $_{[+\varphi]}$  count over surface segments. As in Subsection 4.3.2, this intuition can be formalized through the following two-step definition. First, the constraints are defined for a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose surface string consists of a single segment  $\mathbf{b}$ , as in (129).

$$(129) \quad \begin{aligned} \text{DEP}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} 1 & \text{if } \mathbf{b} \text{ has no underlying correspondent w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \\ 0 & \text{otherwise} \end{cases} \\ \text{UNIF}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} 1 & \text{if } \mathbf{b} \text{ has multiple correspondents w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \\ 0 & \text{otherwise} \end{cases} \\ \text{UNIF}^{\text{grad}}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} k & \text{if } \mathbf{b} \text{ has } k \geq 2 \text{ correspondents w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \\ 0 & \text{otherwise} \end{cases} \\ \text{DEP}_{[+\varphi]}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) &= \begin{cases} 1 & \text{if } \mathbf{b} \text{ has value } + \text{ for feature } \varphi \text{ but } \mathbf{b} \text{ has no correspondent } \mathbf{a} \text{ w.r.t. } \rho_{\mathbf{a}, \mathbf{b}} \\ & \text{such that } \mathbf{a} \text{ also has value } + \text{ for feature } \varphi \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

<sup>6</sup> This implies that the shared correspondence relation must hold between the shared surface string  $\mathbf{b}$  and the “smaller” underlying string  $\mathbf{a}$ .

Second, the constraints are defined for a generic candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  by summing over the segments  $\mathbf{b}$  of the surface string  $\mathbf{b}$ , as in (130). The correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b})}$  which appears on the right-hand side of (130) is the restriction of the original correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  to the surface segment  $\mathbf{b}$ .

$$(130) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{b} \subseteq \mathbf{b}} F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b})})$$

Intuitively, the faithfulness constraint DEPLINEARITY (Heinz 2005) counts over subsequences of length 2 of the surface string. This intuition can be formalized through the following two-step definition. First, the constraint is defined for a candidate  $(\mathbf{a}, \mathbf{b}_1 \mathbf{b}_2, \rho_{\mathbf{a}, \mathbf{b}_1 \mathbf{b}_2})$  whose surface string  $\mathbf{b}_1 \mathbf{b}_2$  has length 2, as in (131a). Second, the constraint is defined for an arbitrary candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  by summing over all surface subsequences  $\mathbf{b}_1 \mathbf{b}_2$  of length 2, as in (131b).

$$(131) \quad \text{a. DEPLIN}(\mathbf{a}, \mathbf{b}_1 \mathbf{b}_2, \rho_{\mathbf{a}, \mathbf{b}_1 \mathbf{b}_2}) = \begin{cases} 1 & \text{if } \mathbf{b}_1, \mathbf{b}_2 \text{ have no correspondents } \mathbf{a}_1, \mathbf{a}_2 \text{ such that } \mathbf{a}_1 \text{ precedes } \mathbf{a}_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{b. DEPLIN}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{b}_1 \mathbf{b}_2 \subseteq \mathbf{b}} \text{DEPLIN}(\mathbf{a}, \mathbf{b}_1 \mathbf{b}_2, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b}_1 \mathbf{b}_2)})$$

Generalizing from these examples, a faithfulness constraint  $F$  is called *additive relative to the surface strings* (O-additive) of order  $\ell$  provided the identity (132) holds for any candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . The sum is over all (possibly overlapping) subsequences of  $\mathbf{b}$  of length  $\ell$ .

$$(132) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{b}_1 \cdots \mathbf{b}_\ell \subseteq \mathbf{b}} F(\mathbf{a}, \mathbf{b}_1 \cdots \mathbf{b}_\ell, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \mathbf{b}_1 \cdots \mathbf{b}_\ell)})$$

$F$  is called *categorical relative to the surface strings* (O-categorical) of order  $\ell$  provided it is additive and furthermore it can only take values 0 or 1 when it is applied to a candidate whose surface string has length  $\ell$ , so that the addenda on the right-hand side of (132) are all equal to 0 or 1. To illustrate, the constraints DEP, UNIFORMITY, and DEP $_{[+\varphi]}$  are O-categorical of order 1; the constraint DEPLINEARITY is O-categorical of order 2; and the constraint UNIFORMITY<sup>grad</sup> is O-additive but not O-categorical.

Finally, a faithfulness constraint  $F$  is called *monotone relative to the surface strings* (O-monotone) provided it satisfies the implication (133): if two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{a}, \widehat{\mathbf{b}}, \rho_{\mathbf{a}, \widehat{\mathbf{b}}})$  share both the underlying string  $\mathbf{a}$  and the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  and only differ because the surface string  $\mathbf{b}$  of the former is a subsequence of the surface string  $\widehat{\mathbf{b}}$  of the latter, then  $F$  assigns less violations to the former candidate than to the latter.

$$(133) \quad \text{If:} \quad \mathbf{b} \subseteq \widehat{\mathbf{b}} \\ \text{Then:} \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) \leq F(\mathbf{a}, \widehat{\mathbf{b}}, \rho_{\mathbf{a}, \widehat{\mathbf{b}}})$$

Any O-additive faithfulness constraint is O-monotone, as a longer surface string yields more addenda on the right-hand side of (132).

#### 4.3.5. McCarthy's (strengthened) categoricity conjecture

McCarthy (2003b) conjectures that every constraint which is relevant for phonological theory is categorical. Given the preceding discussion, this conjecture takes the following form for the faithfulness constraints.

**McCarthy's categoricity conjecture.** *Any faithfulness constraint which is relevant for phonological theory is either C-categorical, or I-categorical or O-categorical.*  $\square$

The notion of categoricity builds on the notion of additivity. As remarked in Subsection 4.3.3, there is a slight formal asymmetry between C-additivity on the one hand and I-/O-additivity on the other hand: all three notions of additivity require a restriction of the correspondence relations; yet, only I-/O-additivity (but not C-additivity) requires additional restrictions on the underlying and surface strings. I will now formulate a slightly stronger version of McCarthy's categoricity conjecture, which takes into account this asymmetry.

The constraints MAX, MAX $_{[+\varphi]}$ , INTEGRITY, and MAXLINEARITY are I-categorical and therefore automatically I-monotone, as shown in Subsection 4.3.2. They are also O-monotone: the number of violations does not shrink (actually, it does not change at all) when the surface string is extended with additional segments (while keeping the correspondence relation unchanged). This is illustrated in (134)



for MAX: the additional surface segment [i] in  $\hat{\mathbf{b}}$  does not affect the number of deleted underlying segments (the correspondence relation is the same in the two candidates, as required by O-monotonicity).

$$(134) \quad \text{a. MAX}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) = 1 \qquad \text{b. MAX}(\mathbf{a}, \hat{\mathbf{b}}, \rho_{\mathbf{a},\mathbf{b}}) = 1$$

$$\begin{array}{ccc} \mathbf{a} = & \text{u k a r} & \\ & | \quad \backslash & \\ & \text{k r a} & \\ \mathbf{b} = & & \end{array} \qquad \begin{array}{ccc} \mathbf{a} = & \text{u k a r} & \\ & | \quad \backslash & \\ & \text{k i r a} & \\ \hat{\mathbf{b}} = & & \end{array}$$

Analogously, the constraints DEP, DEP<sub>[+φ]</sub>, UNIFORMITY, and DEPLINEARITY are O-categorical and therefore automatically O-monotone, as shown in Subsection 4.3.4. They are also I-monotone: the number of violations does not shrink (actually, it does not change at all) when the underlying string is extended with additional segments. These considerations suggest that McCarthy’s categoricity conjecture can be strengthened as follows, complementing categoricity with monotonicity. This revised formulation of the categoricity conjecture captures the formal asymmetry between C-categoricity and I-/O-categoricity. This revision is crucial for proposition 15 to hold.

**Strengthened categoricity conjecture.** *Any faithfulness constraint which is relevant for phonological theory is either C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone.*  $\square$

The conjecture makes no assumptions on the monotonicity of C-categorical constraints. Indeed, the C-categorical constraint I-ADJACENCY (O-ADJACENCY) is O-monotone but not I-monotone (I-monotone but not O-monotone), as additional underlying segments can disrupt adjacency and thus reduce violations.<sup>7</sup>

#### 4.3.6. Summary

Informally, a candidate can be split into smaller “sub-candidates” by splitting either its correspondence relation, or its underlying string, or its surface string. A faithfulness constraint is additive provided the number of violations it assigns to a candidate is the sum of the number of violations it assigns to the sub-candidates, as formalized by the three identities (121), (126), and (132). An additive faithfulness constraint is categorical provided it only takes values 0 or 1 when applied to the sub-candidates, thus formalizing McCarthy’s (2003b) notion of categoricity. Finally, additivity is closely related to monotonicity, which says that the number of violations assigned to bigger candidates (namely, those with more sub-candidates) ought to be larger. I will now use these formal preliminaries in order to obtain an interpretation of the sufficient condition for idempotency in OT provided by the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  in terms of the triangle inequality.

### 4.4. OT Idempotency and the triangle inequality

Chapter 3 has looked at idempotency in OT. In particular, section 3.2 has established the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  (67) as a sufficient condition for OT idempotency, repeated in (135).

$$(135) \quad \text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0 \\ \text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$$

The  $\text{FIC}_{\text{comp}}^{\text{OT}}$  (135) looks like a technical condition without an intuitive interpretation. The rest of this section will derive an intuitive metric interpretation of this condition by investigating its relationship with the  $\text{FTI}_{\text{comp}}$  introduced in section 4.1, repeated in (136).

$$(136) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$$

#### 4.4.1. The $\text{FTI}_{\text{comp}}$ is stronger than the $\text{FIC}_{\text{comp}}^{\text{OT}}$ in the general case

The  $\text{FTI}_{\text{comp}}$  entails the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  in the general case. In fact, assume that the antecedent of the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  (135) holds, namely that  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$ . In this case, the  $\text{FTI}_{\text{comp}}$  (136) coincides with the consequent of the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ , which therefore holds true. The reverse entailment fails in the general case, as shown by the counterexample (137). The candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  in (137a) does not violate the faithfulness

<sup>7</sup> The alternative definition of I-ADJACENCY in footnote 5 makes it I-additive. Crucially, it is also O-monotone: adding surface segments can only disrupt surface adjacency and thus increase the number of violations. Analogous considerations hold for O-ADJACENCY.



#### 4.4.4. On the candidate assumptions of proposition 15

Proposition 15 requires all correspondence relations in the candidate set to be one-to-one, namely rules out both coalescence and breaking (deletion and epenthesis are instead of course allowed).<sup>9</sup> This assumption is unavoidable. In fact, a counterexample such as (76), repeated below in (138), shows that the  $\text{FTI}_{\text{comp}}$  fails for core faithfulness constraints when correspondence relations can break an underlying segment into multiple surface segments.

$$(138) \quad \begin{array}{lll} \text{a. } \text{IDENT}_{[\text{nas}]}(\mathbf{b}, \mathbf{c}) = 0 & \text{b. } \text{IDENT}_{[\text{nas}]}(\mathbf{a}, \mathbf{c}) = 2 & \text{c. } \text{IDENT}_{[\text{nas}]}(\mathbf{a}, \mathbf{b}) = 1 \\ \\ \mathbf{b} = \text{ga} & \mathbf{a} = \text{ɲa} & \mathbf{a} = \text{ɲa} \\ \quad /| | & \quad \quad \quad & \quad | | \\ \mathbf{c} = \text{gga} & \mathbf{c} = \text{gga} & \mathbf{b} = \text{ga} \end{array}$$

Furthermore, the counterexample (137) shows that it also fails when correspondence relations can coalesce multiple underlying segments into a single surface segment.

#### 4.4.5. On the constraint assumptions of proposition 15

Proposition 15 also requires I- and O-categorical faithfulness constraints to satisfy O- and I-monotonicity while no monotonicity conditions are imposed on C-categorical faithfulness constraints. The proof provided in Appendix A.2 shows that this asymmetry between I/O-categoricity and C-categoricity stems from the asymmetry in the definitions of additivity discussed in Subsection 4.3.3: in the case of C-additivity (121), the sub-candidates are obtained by restricting only the correspondence relation; in the case of I-/O-additivity (126)/(132), the sub-candidates are obtained by restricting both the underlying/surface string and the correspondence relation.

The following counterexample shows that the additional monotonicity assumption made by Proposition 15 is indeed crucial. Consider the (unreasonable) faithfulness constraint  $F$  defined in (139) through the by now familiar two steps. It is indeed a faithfulness constraint, namely it assigns zero violations to the identity candidate, in compliance with condition (56).

$$(139) \quad \begin{array}{l} \text{a. } F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \begin{cases} 1 & \text{if the underlying segment } \mathbf{a} \text{ has no surface correspondents and the string } \\ & \mathbf{b} \text{ has length 1} \\ 0 & \text{otherwise} \end{cases} \\ \\ \text{b. } F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{a} \subseteq \hat{\mathbf{a}}} F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}, \hat{\mathbf{a}})}) \end{array}$$

This constraint  $F$  is I-categorical (of order  $\ell = 1$ ) by construction. Yet, the counterexample (140) shows that it is not O-monotone. In fact,  $F$  assigns one violation to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in (140a), because the underlying coda /t/ is deleted and the surface string  $\mathbf{b}$  has length 1. O-monotonicity says that the number of violations cannot decrease if the surface string  $\mathbf{b}$  is extended to  $\hat{\mathbf{b}}$  by adding surface segments. This requirement is shown to fail in (140b): the addition of the onset [p] increases the length of the surface string  $\hat{\mathbf{b}}$  to 2 and thus prevents  $F$  from assigning any violations to the candidate  $(\mathbf{a}, \hat{\mathbf{b}}, \rho_{\mathbf{a}, \mathbf{b}})$

$$(140) \quad \begin{array}{ll} \text{a. } F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = 1 & \text{b. } F(\mathbf{a}, \hat{\mathbf{b}}, \rho_{\mathbf{a}, \mathbf{b}}) = 0 \\ \\ \mathbf{a} = \text{at} & \mathbf{a} = \text{at} \\ \quad | & \quad | \\ \mathbf{b} = \text{a} & \hat{\mathbf{b}} = \text{pa} \end{array}$$

<sup>9</sup> A close look at the proof of Proposition 15 shows that, if the faithfulness constraint  $F$  is I-categorical of order  $\ell = 1$ , the one-to-one assumption can be weakened to the assumption that no correspondence relation coalesces any two underlying segments into a single surface segment. If the faithfulness constraint  $F$  is O-categorical of order  $\ell = 1$ , the one-to-one assumption can be weakened to the assumption that no correspondence relation breaks any underlying segment into two surface segments. If the faithfulness constraint  $F$  is instead C-categorical, the one-to-one assumption cannot be weakened, not even in the case  $\ell = 1$ .



Since an HG typology can be larger than an OT typology, stronger conditions are expected to be needed to discipline all the grammars in the former to satisfy idempotency. Indeed, candidate conditions for HG idempotency are stronger than candidate conditions for OT idempotency, as shown in subsection 4.2.2. The constraint conditions instead turn out to be equivalent on the background of McCarthy’s (strengthened) categoricity conjecture. The faithfulness constraints listed by propositions 2-11 in chapter 3 as satisfying the  $FIC_{\text{comp}}^{\text{OT}}$  thus also satisfy the  $FIC_{\text{comp}}^{\text{HG}}$  (under the further assumption that all correspondence relations are one-to-one).

#### 4.4.8. Summary

This section has shown that the sufficient condition for OT idempotency provided by the  $FIC_{\text{comp}}^{\text{OT}}$  admits a metric interpretation: it effectively requires the faithfulness constraints to measure the phonological distance between underlying and surface representations in compliance with the metric triangle inequality. This interpretation holds under (a slightly stronger version of) McCarthy’s (2003b) categoricity conjecture, formalized in section 4.3.

### A. Proofs

Throughout this appendix, I consider four strings  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ , and  $\mathbf{d}$ , whose generic segments are denoted by  $a$ ,  $b$ ,  $c$ ,  $d$ . I use statements such as “for every/some segment  $\mathbf{a}$ ” as a shorthand for “for every/some segment  $\mathbf{a}$  of the string  $\mathbf{a}$ ”, thus leaving the domain of the quantifier implicit.

#### A.1. Proof of proposition 13

**PROPOSITION 13** *Assume that, for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ , the candidate set also contains a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  such that the  $FIC^{\text{HG}}$  repeated in (142) holds for any faithfulness constraint  $F$  in the constraint set.*

(142) *For every choice of the constant  $\xi \geq 0$ :*

$$\begin{aligned} \text{If:} \quad & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) \leq \xi \\ \text{Then:} \quad & F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) + \xi \end{aligned}$$

*Then, the HG grammar corresponding to any weighting of the constraint set is idempotent, no matter what the markedness constraints look like.  $\square$*

**PROOF.** Suppose that the HG grammar  $G_{\theta}$  corresponding to some weighting  $\theta$  fails at the idempotency implication (52) for some candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , as stated in (143):  $G_{\theta}$  maps the underlying form  $\mathbf{a}$  to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , as required by the antecedent of the idempotency implication; but it fails to map the underlying form  $\mathbf{b}$  to the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , as required by the consequent.

(143)  $G_{\theta}$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  if and only if:

- a.  $G_{\theta}(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ ;
- b.  $G_{\theta}(\mathbf{b}) \neq (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ .

Condition (143b) means that the grammar  $G_{\theta}$  maps the underlying form  $\mathbf{b}$  to a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ . This means that either the two strings  $\mathbf{b}$  and  $\mathbf{c}$  differ or else  $\mathbf{b}$  and  $\mathbf{c}$  coincide but the two correspondence relations  $\rho_{\mathbf{b}, \mathbf{c}}$  and  $\mathbb{I}_{\mathbf{b}, \mathbf{b}}$  differ. The latter option is impossible, because the candidate  $(\mathbf{b}, \mathbf{b}, \rho_{\mathbf{b}, \mathbf{b}})$  with  $\rho_{\mathbf{b}, \mathbf{b}} \neq \mathbb{I}_{\mathbf{b}, \mathbf{b}}$  is harmonically bounded by  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ : faithfulness constraints cannot prefer the former candidate, by (56); and markedness constraints cannot distinguish between the two candidates, by (57). The two strings  $\mathbf{b}$  and  $\mathbf{c}$  must therefore differ and condition (143) becomes (144).

(144)  $G_{\theta}$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  if and only if there exists a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  such that:

- a.  $G_{\theta}(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ ;
- b.  $G_{\theta}(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ .

By assumption, the two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  come with a companion candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$ . The “if-and-only-if” statement (144) can thus be weakened to the “if” statement (145). In fact, if the grammar  $G_{\theta}$  maps the underlying form  $\mathbf{a}$  to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  as stated in (144a), the weights  $\theta$  prefer this candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  to the candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$ , as stated in (145a). Furthermore, if the grammar  $G_{\theta}$  maps the underlying form  $\mathbf{b}$  to the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  as stated in (145b), the weights  $\theta$  prefer this candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  to the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , as stated in (145b).

(145) If  $G_\theta$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ , there exists some candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  such that:

- a.  $\theta$  prefers  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ ,
- b.  $\theta$  prefers  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  to  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ .

Condition (145) can be made explicit as in (146) in terms of the number of constraint violations. These sums run over a generic markedness constraint  $M$  with weight  $\theta_M$  and a generic faithfulness constraint  $F$  with weight  $\theta_F$ . The faithfulness constraints do not appear on the right-hand side of (146b) because  $F(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}}) = 0$  for every faithfulness constraint  $F$ , by (56).

(146) If  $G_\theta$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ , there exists some candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  such that:

- a. 
$$\sum_M \theta_M M(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \sum_F \theta_F F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) < \sum_M \theta_M M(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$$
- b. 
$$\sum_M \theta_M M(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \sum_M \theta_M M(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$$

Let the constant  $\xi$  be defined as  $\xi = \sum_M \theta_M M(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}}) - \sum_M \theta_M M(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ . Since markedness constraints are blind to underlying forms by (57), then also  $\xi = \sum_M \theta_M M(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) - \sum_M \theta_M M(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ . Condition (146) thus becomes:

(147) If  $G_\theta$  fails at idempotency on a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ , there exists some candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  with  $\mathbf{b} \neq \mathbf{c}$  such that:

- a. 
$$\sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) > \sum_F \theta_F F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \xi$$
- b. 
$$\sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \xi$$

In conclusion, idempotency holds for the HG grammar corresponding to any weighting of the constraint set provided the two conditions (147a) and (147b) can never be satisfied both, no matter the choice of the weights  $\theta_F$  and the constant  $\xi$ . In other words, it suffices to assume that, for every two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , there exists some candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  such that:

(148) For every choice of the constant  $\xi \geq 0$ , for every choice of the weights  $\theta_F$ :

- If:** 
$$\sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \xi$$
- Then:** 
$$\sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq \sum_F \theta_F F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \xi$$

To conclude the proof, I need to show that (148) is equivalent to the  $\text{FIC}^{\text{HG}}$  (142). To start, let me show that (142) entails (148). In fact, suppose that the antecedent of the implication (148) holds. For every faithfulness constraint  $F$ , let  $\xi_F$  be defined as in (149a). The antecedent of the implication (148) can thus be restated as in (149b). I can assume without loss of generality that the weights  $\theta_F$  are all different from zero. The position (149a) thus entails (149c). Since the implication (142) holds by hypothesis, (149c) entails (149d). The consequent of the implication (148) thus follows from (149b) by taking the weighted average of the inequalities (149d) over all faithfulness constraints.

- (149) a.  $\xi_F = \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$
- b. 
$$\sum_F \xi_F < \xi$$
- c.  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi_F / \theta_F$
- d.  $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \xi_F / \theta_F$

Let me now show that (148) vice versa entails (142). In fact, suppose that the antecedent of the implication (142) holds, namely that  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi$ . Let me distinguish two cases, depending on whether  $\xi$  is an integer or not. To start, assume that  $\xi$  is not an integer. The assumption  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi$  (with

the loose inequality) is thus equivalent to the assumption  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) < \xi$  (with the strict inequality), because constraint violations are integers. The antecedent of the implication (148) thus holds with all the weights set equal to zero but for the weight  $\theta_F$  corresponding to the faithfulness constraint  $F$  considered, which is equal to 1. The consequent of the implication (148) must therefore hold as well, which is in turn identical to the consequent of the implication (142) with this special choice of the weights. If instead the antecedent of (142) holds with  $\xi$  equal to an integer, let  $\widehat{\xi} = \xi + 1/2$ . By reasoning as above, I conclude that the consequent of the implication (142) holds for  $\widehat{\xi}$ . Since constraint violations are integers, the latter entails in turn that the consequent of the implication (142) holds for  $\xi$ .  $\square$

## A.2. Proof of proposition 15

**PROPOSITION 15** *Assume the candidate set (49) satisfies the transitivity axiom (66) and only contains one-to-one correspondence relations. Consider a faithfulness constraint  $F$  which is C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone.  $F$  satisfies the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  if and only if it satisfies the  $\text{FTI}_{\text{comp}}$ .*  $\square$

**PROOF.** As shown in Subsection 4.4.1, the  $\text{FTI}_{\text{comp}}$  entails the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  in the general case. To prove the reverse entailment, consider a faithfulness constraint  $F$  which satisfies the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  repeated in (150) for any two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  and their composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$ , and let me show that  $F$  then satisfies the  $\text{FTI}_{\text{comp}}$  repeated in (151).

$$(150) \quad \begin{array}{l} \text{If:} \quad F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) = 0 \\ \text{Then:} \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) \end{array}$$

$$(151) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$$

For concreteness, the rest of the proof considers the case where  $F$  is I-categorical of order  $\ell$  and O-monotone, so that it satisfies the I-additivity condition repeated in (152); the cases where  $F$  is instead C-categorical or O-categorical and I-monotone are treated analogously.

$$(152) \quad F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = \sum_{\mathbf{a}_1 \cdots \mathbf{a}_\ell \subseteq \mathbf{a}} F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b})})$$

The I-additivity condition (152) entails that  $F$  assigns zero violations to candidates whose underlying string is shorter than  $\ell$ , as the sum on the right-hand side is empty in this case (there are no subsequences of length  $\ell$ ). The  $\text{FTI}_{\text{comp}}$  (151) thus trivially holds when its string  $\mathbf{a}$  is shorter than  $\ell$ , because its left-hand side  $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$  is equal to zero. From now on, I assume therefore that the string  $\mathbf{a}$  has length at least  $\ell$ .

Consider a subsequence  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  of  $\mathbf{a}$  of length  $\ell$ . Let  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  be the surface correspondent subsequence in  $\mathbf{b}$  of the underlying subsequence  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  relative to the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  (namely,  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  is the subsequence of  $\mathbf{b}$  consisting of all and only the segments which are in correspondence with one of the segments  $\mathbf{a}_1, \dots, \mathbf{a}_\ell$ ). The operations of composition and restriction over correspondence relations commute in the sense of the identity (153): the restriction of the composition correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}}$  to the pair of strings  $(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c})$  coincides with the composition of the restrictions of the relations  $\rho_{\mathbf{a}, \mathbf{b}}$  and  $\rho_{\mathbf{b}, \mathbf{c}}$  to the pairs of strings  $(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})$  and  $(\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c})$ .

$$(153) \quad (\rho_{\mathbf{a}, \mathbf{b}} \circ \rho_{\mathbf{b}, \mathbf{c}}) \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c})} = (\rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})}) \circ (\rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c})})$$

The identity (153) says that the candidate (154c) is the composition of the two candidates (154a) and (154b).

$$(154) \quad \begin{array}{l} \text{a. } (\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})}) \\ \text{b. } (\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c})}) \\ \text{c. } (\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c})}) \end{array}$$

The hypothesis that  $F$  satisfies the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  (150) for these two candidates (154a) and (154b) and their composition candidate (154c) becomes:

$$(155) \quad \begin{array}{l} \text{If:} \quad F(\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c})}) = 0 \\ \text{Then:} \quad F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c})}) \leq \\ \quad \leq F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})}) \end{array}$$

Since  $F$  is I-categorical of order  $\ell$  and since the underlying string  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  has length  $\ell$ , the left-hand side of the inequality in the consequent of (155) is equal to either 0 or 1. By reasoning as in Subsection 4.4.2, the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  (155) thus entails the  $\text{FTI}_{\text{comp}}$  (156).

$$(156) \quad \underbrace{F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c})})}_{(a)} \leq \\ \leq \underbrace{F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})})}_{(b)} + \underbrace{F(\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \mathbf{c})})}_{(c)}$$

The rest of the proof obtains the  $\text{FTI}_{\text{comp}}$  (151) by summing the inequality (156) over all subsequences  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  of length  $\ell$  of the underlying string  $\mathbf{a}$ .

To start, the definition of I-additivity of order  $\ell$  applied to the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$  immediately yields the expression (157) for the sum of the terms (156a) over all subsequences  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  of the underlying string  $\mathbf{a}$ .

$$(157) \quad \sum_{\mathbf{a}_1 \cdots \mathbf{a}_\ell \subseteq \mathbf{a}} \underbrace{F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{c})})}_{(156a)} = F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$$

The sum of the terms (156b) over all subsequences  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  of the underlying string  $\mathbf{a}$  can be upper bounded as in (158). In step (158a), I have used the hypothesis that  $F$  is O-monotone (together with the obvious fact that  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  is a subsequence of  $\mathbf{b}$ ). Step (158b) follows from the fact that the restriction of  $\rho_{\mathbf{a}, \mathbf{b}}$  to the string pair  $(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})$  is identical to its restriction to the string pair  $(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b})$ , because  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  is the subsequence of  $\mathbf{b}$  consisting of those segments which are in correspondence with one of the segments  $\mathbf{a}_1, \dots, \mathbf{a}_\ell$ . Step (158c) follows again from the hypothesis that  $F$  is I-additive of order  $\ell$ .

$$(158) \quad \sum_{\mathbf{a}_1 \cdots \mathbf{a}_\ell \subseteq \mathbf{a}} \underbrace{F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})})}_{(156b)} \leq \\ \stackrel{(a)}{\leq} \sum_{\mathbf{a}_1 \cdots \mathbf{a}_\ell \subseteq \mathbf{a}} F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell})}) \\ \stackrel{(b)}{=} \sum_{\mathbf{a}_1 \cdots \mathbf{a}_\ell \subseteq \mathbf{a}} F(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}} \upharpoonright_{(\mathbf{a}_1 \cdots \mathbf{a}_\ell, \mathbf{b})}) \\ \stackrel{(c)}{=} F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$$

Finally, let me bound the sum of the terms (156c) over all subsequences  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  of the underlying string  $\mathbf{a}$ . To this end, I note that the implication (159) holds for any two subsequences  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  and  $\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell$  and their surface correspondent subsequences  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  and  $\mathbf{b}_{\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell}$  of  $\mathbf{b}$ .

$$(159) \quad \text{If:} \quad \mathbf{a}_1 \cdots \mathbf{a}_\ell \neq \widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell \text{ and } \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell} \text{ has length at least } \ell \\ \text{Then:} \quad \mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell} \neq \mathbf{b}_{\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell}$$

In fact, assume by contradiction that the antecedent holds but the consequent fails. Since the surface correspondent string  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  has length at least  $\ell$  and since the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  cannot break any underlying segment into two or more surface segments (because it is one-to-one), each underlying segment  $\mathbf{a}_i$  of  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  must have a surface correspondent in  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$ . The hypothesis  $\mathbf{a}_1 \cdots \mathbf{a}_\ell \neq \widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell$  means that there exists at least one segment  $\mathbf{a}_i$  which belongs to  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  but not to  $\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell$ . Let  $\mathbf{b}$  be the surface correspondent of  $\mathbf{a}_i$  in  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$ . Because of the contradictory assumption that the consequent of (159) fails,  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell} = \mathbf{b}_{\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell}$ . This means that  $\mathbf{b}$  also belongs to  $\mathbf{b}_{\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell}$ , namely must correspond to some segment  $\widehat{\mathbf{a}}_j$  of  $\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell$ . Since  $\mathbf{a}_i$  does not belong to  $\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_\ell$ , then  $\mathbf{a}_i$  and  $\widehat{\mathbf{a}}_j$  must be different. The conclusion that both  $(\mathbf{a}_i, \mathbf{b})$  and  $(\widehat{\mathbf{a}}_j, \mathbf{b})$  belong to  $\rho_{\mathbf{a}, \mathbf{b}}$  despite the fact that  $\mathbf{a}_i \neq \widehat{\mathbf{a}}_j$  contradicts the hypothesis that  $\rho_{\mathbf{a}, \mathbf{b}}$  does not coalesce any two underlying segments.

Let me now go back to the goal of bounding the sum of the terms (156c) over all subsequences  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  of the underlying string  $\mathbf{a}$ . Since  $\mathbf{a}_1 \cdots \mathbf{a}_\ell$  has length  $\ell$  and since the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  cannot break any underlying segment (because it is one-to-one), the surface correspondent string  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  has length  $\ell$  or smaller. If  $\mathbf{b}_{\mathbf{a}_1 \cdots \mathbf{a}_\ell}$  has length smaller than  $\ell$ , the corresponding term (156c) is null, because  $F$  is I-additive of order  $\ell$  and thus assigns zero violations to candidates whose underlying string is shorter than  $\ell$ , as noted at the beginning. The sum can thus be restricted to candidates whose underlying



form  $\mathbf{b}_{a_1 \dots a_\ell}$  has length exactly  $\ell$ , as in step (160a). Condition (159) says that the mapping from the subsequences  $a_1 \dots a_\ell$  to the corresponding surface subsequences  $\mathbf{b}_{a_1 \dots a_\ell}$  (of length  $\ell$ ) is an injection, thus guaranteeing step (160b). Step (160c) follows again from the hypothesis that  $F$  is I-additive of order  $\ell$ .

$$\begin{aligned}
 (160) \quad & \sum_{\mathbf{a}_1 \dots \mathbf{a}_\ell \subseteq \mathbf{a}} \underbrace{F(\mathbf{b}_{\mathbf{a}_1 \dots \mathbf{a}_\ell}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{b}_{\mathbf{a}_1 \dots \mathbf{a}_\ell}, \mathbf{c})})}_{(156c)} \\
 & \stackrel{(a)}{=} \sum_{\substack{\mathbf{a}_1 \dots \mathbf{a}_\ell \subseteq \mathbf{a} \\ \text{s.t. } \mathbf{b}_{\mathbf{a}_1 \dots \mathbf{a}_\ell} \text{ has length } \ell}} F(\mathbf{b}_{\mathbf{a}_1 \dots \mathbf{a}_\ell}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{b}_{\mathbf{a}_1 \dots \mathbf{a}_\ell}, \mathbf{c})}) \\
 & \stackrel{(b)}{\leq} \sum_{\mathbf{b}_1 \dots \mathbf{b}_\ell \subseteq \mathbf{b}} F(\mathbf{b}_1 \dots \mathbf{b}_\ell, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}} \upharpoonright_{(\mathbf{b}_1 \dots \mathbf{b}_\ell, \mathbf{c})}) \\
 & \stackrel{(c)}{=} F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})
 \end{aligned}$$

The  $\text{FTI}_{\text{comp}}$  (151) follows by summing the inequality (156) over all subsequences  $\mathbf{a}_1 \dots \mathbf{a}_\ell$  of length  $\ell$  of the string  $\mathbf{a}$ , using the three expressions (157), (158), and (160) for the sums over the three terms (156a), (156b), and (156c).  $\square$

## The theory of output-drivenness from the perspective of the triangle inequality

Tesar’s (2013) notion of *output-drivenness* is a formal condition on phonological grammars, construed as mappings from underlying to surface (or *output*) forms. It demands that a grammar which maps an underlying form  $UR_1$  to some surface form  $SR$  also maps to that surface form  $SR$  any other underlying form  $UR_2$  such that  $UR_2$  is more similar to  $SR$  than  $UR_1$  is. The notion of output-drivenness is relevant to phonological theory, because non-output-drivenness unifies various opaque phonological phenomena such as *chain shifts* (Łubowicz 2011) and *derived environment effects* (or *saltations*; White 2014), as we will see in more detail in section 6.2. Furthermore, Tesar shows that output-drivenness has significant learnability implications in the context of the classical *inconsistency detection* approach (Merchant 2008) to the problem of learning a lexicon of underlying forms from a paradigm of surface forms.

The main result of Tesar’s theory guarantees output-drivenness of an entire typology of OT grammars when the correspondence relations are one-to-one and the constraint set only contains three types of faithfulness constraints, namely DEP, MAX, and IDENT. This chapter reconstructs Tesar’s theory from the perspective of the faithfulness triangle inequality formulated in section 4.1. To highlight the role of the triangle inequality, I define the notion of phonological similarity that output-drivenness is predicated on *axiomatically* through a condition on the faithfulness constraints, rather than concretely in terms of strings and correspondence relations, as Tesar does. This alternative approach yields a stronger result which ensures a slightly stronger notion of output-drivenness for a larger variety of faithfulness constraints, beyond the three faithfulness constraints considered by Tesar.

This chapter is organized as follows. Section 5.1 recalls Tesar’s notion of output-drivenness, highlighting its dependence on an underlying notion of phonological similarity. Section 5.2 recalls Tesar’s result that the output-drivenness of the grammars in an OT typology is guaranteed by a condition on the faithfulness constraints, called the *faithfulness output-drivenness condition* (FODC). Section 5.3 reviews in detail Tesar’s concrete definition of the notion of phonological similarity that output-drivenness is predicted on and section 5.4 introduces my axiomatization thereof, in terms of a condition on a selected set of faithfulness constraints (which could in principle be smaller than or different from the faithfulness constraint set used to define the typology). Section 5.5 presents the main result of this chapter: although the FODC has an apparently very technical nature, it is actually equivalent to the faithfulness triangle inequality when phonological similarity is defined in terms of the proposed axiomatization. This result in turns entails an equivalence between the sufficient condition for idempotency provided by the FIC and the sufficient condition for output-drivenness provided by the FODC, for those faithfulness constraints which are categorical and furthermore figure in the faithfulness constraint set used to measure phonological similarity. This observation strengthens Tesar’s result on the output-drivenness of OT grammars: from the case where the constraint set only contains the three faithfulness constraints DEP, MAX, and IDENT to the case where it contains any of the faithfulness constraints considered in chapter 3.

### 5.1. Output-drivenness and its relationship with idempotency

This section reviews Tesar’s notion of output-drivenness and compares it to the notion of idempotency considered in the receding chapters.

#### 5.1.1. Tesar’s notion of output-drivenness

Consider two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  which share a surface form  $\mathbf{d}$ . Suppose that the underlying form  $\mathbf{b}$  is more similar to  $\mathbf{d}$  than the other underlying form  $\mathbf{a}$  is. In other words, that the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  has more *internal similarity* than the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$ . Tesar formalizes this assumption through the condition (161), where  $\leq_{\text{sim}}$  is a *similarity order*, namely an ordering relation among candidates (or, more precisely, among candidates which share the surface form) based on their

internal similarity. Subsections 5.3-5.4 will address the issue of the proper definition of this similarity order.

$$(161) \quad (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$$

Suppose that a phonological grammar  $G$  maps the less similar underlying form  $\mathbf{a}$  to the surface form  $\mathbf{d}$ , namely that  $G(\mathbf{a}) = (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ . This means that  $\mathbf{d}$  is phonotactically licit and that  $\mathbf{a}$  is not too dissimilar from  $\mathbf{d}$ . Since the phonotactic status of  $\mathbf{d}$  does not depend on the underlying form and furthermore  $\mathbf{b}$  is even more similar to  $\mathbf{d}$ , the grammar  $G$  should map also the more similar underlying form  $\mathbf{b}$  to that same surface form  $\mathbf{d}$ , namely  $G(\mathbf{b}) = (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ . A grammar which abides by this logics is called *output-driven*.

**Definition 2.** [Tesar's Output-drivenness] A grammar  $G$  is *output-driven* relative to a similarity order  $\leq_{\text{sim}}$  provided the following implication holds

$$(162) \quad \begin{array}{ll} \text{If:} & G(\mathbf{a}) = (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) & (\text{candidate with less similarity}) \\ \text{Then:} & G(\mathbf{b}) = (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) & (\text{candidate with more similarity}) \end{array}$$

for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  which share the surface form  $\mathbf{d}$  and satisfy the condition  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ .  $\square$

### 5.1.2. Output-drivenness entails idempotency

Intuitively, any string  $\mathbf{b}$  is more similar to itself than to any other string  $\mathbf{a}$ . In other words, identity candidates have the greatest internal similarity. A similarity order  $\leq_{\text{sim}}$  thus needs to satisfy condition (163) for any candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and the corresponding identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ , whose existence is guaranteed by the reflexivity axiom (51).

$$(163) \quad (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$$

Whenever condition (163) holds, output-drivenness entails idempotency. In fact, output-drivenness requires the implication (162) to hold for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that the former has less internal similarity than the latter. Condition (163) ensures that is indeed the case when the two strings  $\mathbf{b}$  and  $\mathbf{d}$  coincide and  $\rho_{\mathbf{b},\mathbf{d}}$  is the identity on the string  $\mathbf{b} = \mathbf{d}$ . In this case, the implication (162) in the definition of output-drivenness specializes to (164), which is in turn the implication (52) in the definition of idempotency.

$$(164) \quad \begin{array}{ll} \text{If:} & G(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \\ \text{Then:} & G(\mathbf{b}) = (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}}) \end{array}$$

In conclusion, the definition of idempotency coincides with the definition of output-drivenness in the special case where  $\mathbf{b} = \mathbf{d}$  and  $\rho_{\mathbf{b},\mathbf{d}}$  is the identity.

### 5.1.3. Output-drivenness is stronger than idempotency: saltations

Although output-drivenness entails idempotency, the reverse entailment fails: output-drivenness is a stronger condition than idempotency. As a counterexample, consider the *derived environment effect* or *saltation* (Łubowicz 2002, White 2013) in (165) from the Campidanian dialect of Sardinia: voiceless stops are lenited to voiced fricatives in post-vocalic position, while voiced stops are faithfully realized (Bolognesi 1998, via White 2013).

(165)	<i>isolated form</i>	<i>post-vocalic form</i>	
	/p, t, k/ → [β, ð, ɣ]:	[piʃ:i]	[bel:u βiʃ:i] ‘(nice) fish’
		[trintaduzu]	[s:u ðrintaduzu] ‘(the) thirty-two’
		[kuat:ru]	[dɛ ɣuat:ru] ‘(of) four’
	/b, d, g/ → [b, d, g]:	[bĩu]	[s:u bĩu] ‘(the) wine’
		[dominiyu]	[don:ja dominiyu] ‘(every) Sunday’
		[gɔma]	[dɛ gɔma] ‘(of) rubber’

A reasonable definition of the similarity order  $\leq_{\text{sim}}$  (see below footnote 4) should guarantee that the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  in (166) has less internal similarity than the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , since the former involves a disparity for both voicing and continuancy, while the latter only for continuancy.

$$(166) \quad \begin{array}{ccc} \mathbf{a} = /p/ & & \mathbf{b} = /b/ \\ | & & | \\ \mathbf{d} = [\beta] & & \mathbf{d} = [\beta] \end{array}$$

The phonological pattern (165) can thus be represented as in (167): the voiced stop  $\mathbf{a}$  is mapped to the voiced fricative  $\mathbf{d}$ , skipping or jumping over the closer licit voiced stop  $\mathbf{b}$ .

$$(167) \quad \begin{array}{ccc} & \curvearrowright & \\ \mathbf{a} = p & & \mathbf{d} = \beta \\ & \curvearrowleft & \\ & \mathbf{b} = b & \end{array}$$

Since  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ , output-drivenness would require the more similar underlying form  $\mathbf{b} = /b/$  to be mapped to the surface form  $\mathbf{d} = [\beta]$  when the less similar underlying form  $\mathbf{a} = /p/$  is mapped to  $\mathbf{d} = [\beta]$ . The saltation pattern (165) thus fails at output-drivenness, despite being idempotent.

## 5.2. Output-drivenness in OT

Which conditions on the candidate and constraint sets ensure that the OT grammars corresponding to any ranking is output-driven? This section reviews Tesar's answer to this question.

### 5.2.1. A condition on the candidate set: the one-to-one assumption

Let me start with candidate conditions. As seen in sections 3.3.2 and 3.3.3, OT idempotency fails when correspondence relations can break an underlying segment into multiple surface segments. Since output-drivenness entails idempotency, the no-breaking assumption is needed for output-drivenness as well. The situation is different for coalescence of two underlying segments into a single surface segment. While coalescence does not threaten OT idempotency, it does hinder output-drivenness, as shown by the following counterexample. A reasonable definition of the similarity order  $\leq_{\text{sim}}$  (see below footnote 4) should guarantee that the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  in (168), with coalescence of the underlying complex coda, has less internal similarity than the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ .

$$(168) \quad \begin{array}{ccc} \mathbf{a} = \text{tetd} & & \mathbf{b} = \text{tad} \\ \begin{array}{c} ||| \backslash \\ \mathbf{d} = \text{tat} \end{array} & & \begin{array}{c} ||| \\ \mathbf{d} = \text{tat} \end{array} \end{array}$$

Although  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ , the constraint  $M = * \text{VOICE}$  in the grammar (169) maps the underlying form  $\mathbf{a}$  to the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  with less internal similarity; yet, it fails at mapping the underlying form  $\mathbf{b}$  to the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  with more internal similarity, violating output-drivenness.

(169)	/tetd/	IDENT <sub>[vce]</sub>	M	/tad/	IDENT <sub>[vce]</sub>	M
☞	$\mathbf{a} = \text{tetd}$     \	*		$\mathbf{b} = \text{tad}$ 	*!	
	$\mathbf{d} = \text{tat}$			$\mathbf{d} = \text{tat}$		
	$\mathbf{a} = \text{tetd}$     \	*	*!	☞ $\mathbf{b} = \text{tad}$ 		*
	$\mathbf{c} = \text{tad}$			$\mathbf{c} = \text{tad}$		

In conclusion, output-drivenness fails even in the simplest cases when correspondence relations display breaking or coalescence and thus are not one-to-one.

### 5.2.2. A condition on the faithfulness constraint set: the FODC<sup>OT</sup>

The following Proposition 18 ensures output-drivenness when all the faithfulness constraints satisfy the two implications (170). These implications are thus jointly referred to as the OT *faithfulness output-drivenness condition* (FODC<sup>OT</sup>). No assumptions are made on the markedness constraints, on the nature of the faithfulness constraints (for instance, they are not required to be categorical), on the correspondence relations in the candidate set (for instance, they are not required to be one-to-one), or on the similarity order (which is left completely unspecified). Proposition 18 was derived in Tesar (2013, chapter 3) and its proof is recalled in appendix A.1 for completeness. This proposition 18 is analogous to Proposition 1 in Subsection 3.2.3, which was indeed derived by mimicking Tesar's reasoning.

**Proposition 18.** *Assume that, for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , for any other candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , the candidate set also contains a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  different from  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  such that the two following implications (170) hold for every faithfulness constraint  $F$  in the constraint set.*

- (170) a. **If:**  $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$   
**Then:**  $F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$   
b. **If:**  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$   
**Then:**  $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) < F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$

Then, the OT grammar corresponding to any ranking of the constraint set is output-driven relative to the similarity order  $\leq_{sim}$ .  $\square$

### 5.2.3. The FODC<sup>OT</sup> entails the FIC<sup>OT</sup>

Subsection 5.1.2 has shown that output-drivenness entails idempotency because the definition of idempotency coincides with the definition of output-drivenness in the special case of maximal similarity where the two strings  $\mathbf{b}$  and  $\mathbf{d}$  coincide. This entailment relationship carries over to the two sufficient conditions for output-drivenness and idempotency: the FODC<sup>OT</sup> entails the FIC<sup>OT</sup> derived in subsection 3.2.3. In fact, condition (163) on the similarity order ensures that the FODC<sup>OT</sup> holds in the special case where the two strings  $\mathbf{b}$  and  $\mathbf{d}$  coincide (and the correspondence relation  $\rho_{\mathbf{b},\mathbf{d}}$  is the identity over the string  $\mathbf{b} = \mathbf{d}$ ). In this special case, the contrapositive of the first FODC<sup>OT</sup> implication (170a) specializes to (171).

- (171) **If:**  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$   
**Then:**  $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$

By condition (56) in subsection 3.2.1, faithfulness constraints assign no violations to identity candidates. Since violation numbers are non-negative, the antecedent of (171) is equivalent to the condition  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$ . The (contrapositive of) the first FODC<sup>OT</sup> implication in (171) thus coincides with the FIC<sup>OT</sup> (65) from subsection 3.2.3, repeated in (172).<sup>1</sup>

- (172) **If:**  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$   
**Then:**  $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$

### 5.2.4. The FODC<sup>OT</sup> is stronger than the FIC<sup>OT</sup>

Subsection 5.1.3 has shown that output-drivenness is a stronger condition on phonological grammars than idempotency, because output-drivenness also excludes idempotent phonological patterns such as saltations. This relationship carries over to the two sufficient conditions for output-drivenness and idempotency: the FODC<sup>OT</sup> is stronger than the FIC<sup>OT</sup> repeated above in (172). Here is an interesting counterexample which illustrates this fact. Consider the faithfulness constraint  $F = \text{IDENT}_{[\text{voice}]} \vee \text{IDENT}_{[\text{cont}]}$  which is the disjunction of the two identity faithfulness constraints  $\text{IDENT}_{[\text{voice}]}$  and  $\text{IDENT}_{[\text{cont}]}$  for voicing and continuancy (Downing 1998, 2000). Subsection 3.5.6 has shown that  $F$  satisfies the FIC<sup>OT</sup> (provided the correspondence relations do not break any underlying segment). Let me show that  $F$  nonetheless does not satisfy the FODC<sup>OT</sup>. Consider again the candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  in (166), repeated in (173). As noted above, any reasonable definition of the similarity order  $\leq_{sim}$  should guarantee that the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  has less internal similarity than the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , since the former involves a disparity for two features (voicing and continuancy) while the latter involves a disparity for only one feature (continuancy, not voicing). Thus,  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ .

- (173) a.  $\mathbf{a} = /p/$                       b.  $\mathbf{b} = /b/$   
          |                                      |  
           $\mathbf{d} = [\beta]$                           $\mathbf{d} = [\beta]$

Consider  $\mathbf{b} = \mathbf{c}$ . The antecedent of the second FODC<sup>OT</sup> implication (170b) holds, as shown in (174a). Yet, its consequent fails, as shown in (174b).

<sup>1</sup> The second FODC<sup>OT</sup> implication (170b) is trivially satisfied in the special case where  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) = (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ , because its antecedent  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  becomes  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < 0$ , which contradicts the non-negativity of constraint violations.



**Deletion clause:** there exists an injection from the segments of **b** deleted relative to  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  to the segments of **a** deleted relative to  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$ .<sup>2</sup>

**Epenthesis clause:** every segment of **d** which is epenthetic relative to  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  is also epenthetic relative to  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$ .  $\square$

Let me illustrate Definition 3 with a couple of examples which will be relevant for what follows. The two candidates in (176a) satisfy the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\Phi, \text{weak}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ . In fact, the epenthesis clause trivially holds because neither candidate displays epenthesis. The identity clause holds relative to the feature set  $\Phi = \{[\text{high}], [\text{voice}]\}$ , because the more similar candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  displays no feature mismatches. Finally, the deletion clause holds because both candidates feature exactly one deleted coda. Note crucially that the two deleted underlying codas differ from each other in voicing.

$$(176) \quad \text{a.} \quad \begin{array}{cc} \mathbf{a} = \text{r i d} & \mathbf{b} = \text{r e t} \\ \quad \quad \quad \parallel & \quad \quad \quad \parallel \\ \mathbf{d} = \text{r e} & \mathbf{d} = \text{r e} \end{array} \quad \text{b.} \quad \begin{array}{cc} \mathbf{a} = \text{p a} & \mathbf{b} = \text{k i} \\ \quad \quad \quad \parallel & \quad \quad \quad \parallel \\ \mathbf{d} = \text{t i} & \mathbf{d} = \text{t i} \end{array}$$

As another example, the two candidates in (176b) satisfy the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\Phi, \text{weak}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ . In fact, the epenthesis and deletion clauses trivially hold, because neither candidate features epenthesis or deletion. Consider the feature set  $\Phi = \{[\text{high}], [\text{place}]\}$  and assume that [place] is a three-valued feature. The identity clause holds because the underlying onset of both candidates has a surface correspondent which mismatches in place but only the underlying vowel of the less similar candidate has a surface correspondent which mismatches in height. Note crucially that the two underlying onsets differ from each other in place of articulation.

### 5.3.2. Towards a stronger similarity order

Unfortunately, output-drivenness fails even in the simplest cases relative to the similarity order provided by Definition 3. To start, consider the two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  in (176a). Suppose that, because of a higher ranked markedness constraint against voiced codas (omitted here), the OT grammar (177a) maps the underlying form **a** to the surface form **d** (which deletes the voiced coda) rather than to **c** (which instead devoices it). Since  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ , output-drivenness requires that grammar to map also the underlying form **b** to **d**, contrary to (177b).<sup>3</sup>

$$(177) \quad \text{a.} \quad \begin{array}{|c|c|c|} \hline & /rid/ & \text{ID}_{[\text{voice}]} \text{ MAX} \\ \hline \mathbf{a} = \text{r i d} & & *! \\ \quad \quad \quad \parallel & & \\ \mathbf{c} = \text{r e t} & & \\ \hline \mathbb{E} \mathbf{a} = \text{r i d} & & * \\ \quad \quad \quad \parallel & & \\ \mathbf{d} = \text{r e} & & \\ \hline \end{array} \quad \text{b.} \quad \begin{array}{|c|c|c|} \hline & /ret/ & \text{ID}_{[\text{voice}]} \text{ MAX} \\ \hline \mathbb{E} \mathbf{b} = \text{r e t} & & \\ \quad \quad \quad \parallel & & \\ \mathbf{c} = \text{r e t} & & \\ \hline \mathbf{b} = \text{r e t} & & * \\ \quad \quad \quad \parallel & & \\ \mathbf{d} = \text{r e} & & \\ \hline \end{array}$$

As another counterexample, consider the two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  in (176b). The OT grammar (178a) maps the underlying form **a** to the surface form **d**. Since  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ , output-drivenness requires that grammar to map also the underlying form **b** to **d**, contrary to (177b).

$$(178) \quad \text{a.} \quad \begin{array}{|c|c|c|} \hline & /pa/ & \text{ID}_{[\text{place}]} \text{ *}[k] \\ \hline \mathbf{a} = \text{p a} & * & *! \\ \quad \quad \quad \parallel & & \\ \mathbf{b} = \text{k i} & & \\ \hline \mathbb{E} \mathbf{a} = \text{p a} & * & \\ \quad \quad \quad \parallel & & \\ \mathbf{d} = \text{t i} & & \\ \hline \end{array} \quad \text{b.} \quad \begin{array}{|c|c|c|} \hline & /ki/ & \text{ID}_{[\text{place}]} \text{ *}[k] \\ \hline \mathbb{E} \mathbf{b} = \text{k i} & & * \\ \quad \quad \quad \parallel & & \\ \mathbf{b} = \text{k i} & & \\ \hline \mathbf{b} = \text{k i} & *! & \\ \quad \quad \quad \parallel & & \\ \mathbf{d} = \text{t i} & & \\ \hline \end{array}$$

<sup>2</sup> This means that there exists a mapping from the segments of **b** deleted relative to  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  to the segments of **a** deleted relative to  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  such that two different deleted segments of **b** correspond to two different deleted segments of **a**. In other words,  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  has at least as many deleted segments as  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ .

<sup>3</sup> It is an interesting exercise to check that this failure of output-drivenness comes, as expected, with a failure of the FODC<sup>OT</sup>. Indeed, the first FODC<sup>OT</sup> (170a) fails for  $F = \text{IDENT}_{[\text{voice}]}$ :  $\text{ID}(\mathbf{a}, \mathbf{d}) = 0$  and  $\text{ID}(\mathbf{a}, \mathbf{c}) = 1$ , so that the antecedent of the first FODC<sup>OT</sup> (170a) holds; yet  $\text{ID}(\mathbf{b}, \mathbf{d}) = \text{ID}(\mathbf{b}, \mathbf{c}) = 0$ , so that its consequent fails. Other choices of the correspondence relation  $\rho_{\mathbf{a}, \mathbf{c}}$  yields analogous failures. For instance, suppose that  $\rho_{\mathbf{a}, \mathbf{c}}$  establishes no correspondence between the codas of **a** and **c**. In this case, the second FODC<sup>OT</sup> (170b) fails for  $F = \text{MAX}$ :  $\text{MAX}(\mathbf{b}, \mathbf{c}) = 0$  and  $\text{MAX}(\mathbf{b}, \mathbf{d}) = 1$ , so that the antecedent of the second FODC<sup>OT</sup> (170b) holds; yet  $\text{MAX}(\mathbf{a}, \mathbf{c}) = \text{MAX}(\mathbf{a}, \mathbf{d}) = 1$ , so that its consequent fails.

These counterexamples show that Definition 3 is too weak: it yields a similarity order which holds between too many pairs of candidates, making output-drivenness too hard to satisfy. We thus need a stronger definition of the similarity order, which is satisfied by less pairs of candidates and in particular is not satisfied by pairs of candidates such as those in (176), which we have just seen to be problematic for output-drivenness.

### 5.3.3. Tesar’s similarity order

Indeed, the similarity order considered by Tesar is not the one provided by Definition 3 above but the one provided by the following Definition 4. The two definitions differ for the material underlined.

**Definition 4.** [Tesar’s similarity order] For any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  sharing a surface form  $\mathbf{d}$ , let  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\Phi} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  provided:

**Identity clause:** for every feature  $\varphi$  in  $\Phi$ , for every segment  $\mathbf{d}$  of  $\mathbf{d}$ , if there exists a segment  $\mathbf{b}$  of  $\mathbf{b}$  such that  $(\mathbf{b}, \mathbf{d}) \in \rho_{\mathbf{b}, \mathbf{d}}$  and  $\varphi(\mathbf{b}) \neq \varphi(\mathbf{d})$ , there exists a segment  $\mathbf{a}$  of  $\mathbf{a}$  such that  $(\mathbf{a}, \mathbf{d}) \in \rho_{\mathbf{a}, \mathbf{d}}$  and  $\varphi(\mathbf{a}) = \varphi(\mathbf{b})$ .

**Deletion clause:** there exists an injection from the segments of  $\mathbf{b}$  deleted relative to  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  to the segments of  $\mathbf{a}$  deleted relative to  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  such that any two deleted segments corresponding through the injection agree on the value of every feature in  $\Phi$ .

**Epenthesis clause:** every segment of  $\mathbf{d}$  which is epenthetic relative to  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  is also epenthetic relative to  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$ .  $\square$

The deletion clause of the new Definition 4 is stronger than the corresponding clause in the original Definition 3 because of the additional boxed condition which requires two deleted segments (corresponding through the injection) to match in feature values. To illustrate, consider the two candidates in (176a). The two deleted underlying codas differ relative to the feature [voice]. The additional underlined condition of the deletion clause therefore fails. The two candidates in (176a) thus fail the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\Phi} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  relative to the similarity order  $\leq_{\text{sim}}^{\Phi}$  provided by the revised Definition 4. The problem with the failure of output-drivenness in (177) is therefore circumvented.

The identity clause of the new Definition 4 has the underlined condition  $\varphi(\mathbf{a}) = \varphi(\mathbf{b})$  instead of the condition  $\varphi(\mathbf{a}) \neq \varphi(\mathbf{d})$  of the original Definition 3. Because of the assumption  $\varphi(\mathbf{b}) \neq \varphi(\mathbf{d})$ , the condition  $\varphi(\mathbf{a}) = \varphi(\mathbf{b})$  of the new Definition 4 entails the condition  $\varphi(\mathbf{a}) \neq \varphi(\mathbf{d})$  of the original Definition 3. The reverse entailment fails whenever the feature  $\varphi$  has more than two values. To illustrate, consider the two candidates in (176b). The underlying onset /k/ of the underlying form  $\mathbf{b}$  differs from its surface correspondent [t] relative to the feature [place]. The underlying onset /p/ of  $\mathbf{a}$  also differs from its surface correspondent [t]. The identity condition of the original Definition 3 thus holds. But the identity condition of the new Definition 4 fails, because the two underlying onsets /p/ and /k/ differ relative to the three-valued feature [place]. The two candidates in (176b) thus fail the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\Phi} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  relative to the similarity order  $\leq_{\text{sim}}^{\Phi}$  provided by the revised Definition 4. The problem with the failure of output-drivenness in (178) is therefore circumvented.

## 5.4. Axiomatizing the similarity order

As discussed in section 5.3, Tesar’s Definition 4 describes the similarity order *concretely*, in terms of strings and correspondence relations. Within constraint-based phonology, it is natural to assess similarity through the faithfulness constraints. This section thus provides an axiomatization of the similarity order in terms of conditions on the faithfulness constraints. The proposed axiomatization subsumes Tesar’s concrete similarity order provided by Definition 4 as a special case. Despite its reliance on faithfulness constraints, I submit that my alternative approach is not inconsistent with the desired framework-independency of output-drivenness. To start, the technical notion of *disparity* that Tesar relies on for its definition of similarity is really nothing else than a different name for a faithfulness constraint violation. Furthermore, Tesar does not shy away from correspondence relations despite the fact that they are a representational device needed in constraint-based phonology to get around the lack of phonological derivations. In fact, Tesar (p. 34) explains that, “while in linguistics the terminology of correspondence is perhaps found most explicitly in the OT literature, the concept is equally important to any generative theory. There is a correspondence relation implicit in every SPE-style rule.” The same argument applies to faithfulness constraints: although they were only formalized in OT, faithfulness considerations are plausibly intrinsic to phonological theorizing, independently of the framework. Finally, Tesar’s concrete definition of similarity is tailored to restrictive representational assumptions, which only allow for



deletion, epenthesis and feature mismatches but no additional faithfulness violations. Indeed, Tesar’s application of output-drivenness to OT presupposes a faithfulness constraint set limited to MAX, DEP and IDENT constraints. An axiomatization of similarity in terms of an arbitrary set of faithfulness constraints seems to me the most straightforward way to generalize Tesar’s theory of output-drivenness to richer representational assumptions, as shown in the next section. I conclude that there is no principled impediment against rephrasing Tesar’s definition of similarity in terms of faithfulness constraints.

#### 5.4.1. An initial attempt

Intuitively, a candidate has more internal similarity than another candidate provided the former incurs less faithfulness violations than the latter. This intuition is formalized by Definition 5. For the sake of generality (see below Subsection 5.5.4), the similarity order is parameterized by a set  $\mathcal{F}$  of faithfulness constraints which can in principle be different from or smaller than the entire faithfulness constraint set used to define the OT typology.

**Definition 5.** For any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  sharing the surface form  $\mathbf{d}$ , let  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}, \text{weak}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  provided the following inequality holds

$$(179) \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$$

for every faithfulness constraint  $F$  in the faithfulness constraint set  $\mathcal{F}$ .  $\square$

When the faithfulness constraint set  $\mathcal{F}$  consists of MAX, DEP, and  $\text{IDENT}_{\varphi}$  corresponding to features  $\varphi \in \Phi$ , the similarity order  $\leq_{\text{sim}}^{\mathcal{F}, \text{weak}}$  provided by the faithfulness-based Definition 5 coincides with the weak similarity order  $\leq_{\text{sim}}^{\Phi, \text{weak}}$  provided by the concrete Definition 3 and thus inherits the drawbacks discussed in Subsection 5.3.2. A stronger axiomatization is therefore needed.

#### 5.4.2. A stronger axiomatization

Subsection 5.2.5 has shown that the sufficient condition for output-drivenness provided by the  $\text{FODC}^{\text{OT}}$  (170) is tight: any faithfulness constraint which fails the  $\text{FODC}^{\text{OT}}$  yields a failure of output-drivenness. In order to amend Definition 5, I thus look at the  $\text{FODC}^{\text{OT}}$ . Given two candidates such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , Tesar’s proposition 18 requires the  $\text{FODC}^{\text{OT}}$  to hold for *every* candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  and *some* candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ . Consider the special case where  $\mathbf{c} = \mathbf{b}$  and  $\rho_{\mathbf{b},\mathbf{c}}$  is the identity  $\mathbb{I}_{\mathbf{b},\mathbf{b}}$  over the string  $\mathbf{b} = \mathbf{c}$ . Tesar’s proposition thus requires in particular that there exists *some* candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  such that the first  $\text{FODC}^{\text{OT}}$  implication (170a) holds, which in the specific case considered becomes (180).

$$(180) \quad \begin{array}{ll} \text{If:} & F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \\ \text{Then:} & F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}}) \end{array}$$

By (56), faithfulness constraints assign zero violations to the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ . The consequent of the implication (180) thus says that  $F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < 0$ . The latter contradicts the assumption that constraint violations are non-negative. Since the consequent of the implication (180) is false, the antecedent must be false as well. This means that the similarity order  $\leq_{\text{sim}}$  must be defined in such a way that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  entails the existence of some candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which satisfies the inequality (181).

$$(181) \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$$

In conclusion, the similarity order  $\leq_{\text{sim}}$  must be defined in such a way that the inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  entails both the intuitive condition (179) and the technical condition (181). The simplest way to achieve that is by adding the right-hand side of those two conditions yielding (182). This revised condition is stronger than the original condition (179) because of the additional term on the right-hand side of the inequality. Intuitively, this new condition (182) requires the less similar candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  to “make up” not only for any faithfulness violation incurred by the more similar candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  but also for any faithfulness violation incurred by the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which puts the two strings  $\mathbf{a}$  and  $\mathbf{b}$  in correspondence.

**Definition 6.** [Axiomatization of similarity orders] For any candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  which share a surface form  $\mathbf{d}$ , let  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  provided the candidate set contains a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which puts in correspondence the two strings  $\mathbf{a}$  and  $\mathbf{b}$  in such a way that the following inequality holds



### 5.4.3. Formal properties of the strong similarity order

The following proposition (proven in appendix A.2) shows that the relation  $\leq_{\text{sim}}^{\mathcal{F}}$  provided by definition 6 is indeed a partial order on the candidate set and that it satisfies the intuitive condition (163) that identity candidates have maximal internal similarity. Note that the reflexivity axiom in (187) is stronger than the reflexivity axiom (51) from section 3.1: the latter only requires the existence of the identity candidate relative to the surface form, while the former also requires the existence of the identity candidate relative to the underlying form.

**Proposition 19.** *Assume the candidate set (49) satisfies the following reflexivity axiom (187) as well as the transitivity axiom (66) of section 3.2, repeated in (188)*

(187) *If the candidate set contains a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ , it also contains the corresponding identity candidates  $(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a},\mathbf{a}})$  and  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$ , where  $\mathbb{I}_{\mathbf{a},\mathbf{a}}$  and  $\mathbb{I}_{\mathbf{b},\mathbf{b}}$  are the identity correspondence relations among the segments of  $\mathbf{a}$  and  $\mathbf{b}$ , respectively.*

(188) *If the candidate set contains two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which share  $\mathbf{b}$  as the surface and the underlying form, it also contains the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  whose correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  is the composition of  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$ .*

*Assume that the faithfulness constraint subset  $\mathcal{F}$  is complete: for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ ,  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  which share the surface form, there exists a faithfulness constraint  $F$  in  $\mathcal{F}$  which assigns them a different number of violations. Assume furthermore that each faithfulness constraint  $F$  in  $\mathcal{F}$  satisfies the triangle inequality  $F\text{TI}_{\text{comp}}$  repeated in (189) from section 4.1.*

$$(189) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$$

*Then, the relation  $\leq_{\text{sim}}^{\mathcal{F}}$  provided by definition 6 is a partial order on the candidate set and it satisfies the condition (163) that identity candidates have maximal internal similarity.  $\square$*

### 5.4.4. The axiomatic definition subsumes Tesar's concrete definition

Assume that the faithfulness constraint set  $\mathcal{F}$  consists of the faithfulness constraints that Tesar focuses on, namely MAX, DEP, and  $\text{IDENT}_{\varphi}$  for any feature  $\varphi$  in the feature set  $\Phi$ . The sophisticated analysis developed in Tesar (2013, chapter 3) can be rebooted (see Magri 2016b for details) to establish that:

**Proposition 20.** *For any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , if the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\Phi} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  holds relative to the similarity order  $\leq_{\text{sim}}^{\Phi}$  provided by Tesar's concrete Definition 4, then the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  holds relative to the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  provided by the axiomatic Definition 6.  $\square$*

Yet, the two similarity orders  $\leq_{\text{sim}}^{\Phi}$  and  $\leq_{\text{sim}}^{\mathcal{F}}$  are not equivalent: it is easy to construct cases where  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  but  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \not\leq_{\text{sim}}^{\Phi} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , even when  $\mathcal{F}$  only consists of the three types of faithfulness constraints that Tesar focuses on (see Magri 2016b for discussion). In conclusion, the proposed axiomatization of the similarity order subsumes Tesar's concrete definition as a special, concrete case. The notion of output-drivenness relative to the axiomatized similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  is thus slightly stronger than Tesar's original notion (because it holds relative to a similarity order which is slightly looser) and the resulting theory therefore slightly more general.

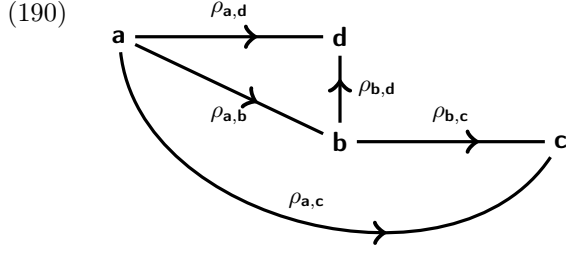
### 5.4.5. The composition candidate and the $\text{FODC}_{\text{comp}}^{\text{OT}}$

Consider two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  which satisfy the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ . Tesar's Proposition 18 considers an arbitrary third candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which puts the more similar underlying form  $\mathbf{b}$  in correspondence with a candidate  $\mathbf{c}$  different from  $\mathbf{d}$  through some relation  $\rho_{\mathbf{b},\mathbf{c}}$ . The proposition then requires the existence of a fourth candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  which puts the less

Consider next the candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  in (168), repeated below in (iia) and (iib). In order to secure the argument made in Subsection 5.2.1, we need to secure the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  with  $\mathcal{F} = \{\text{IDENT}_{[\text{voice}]}, \text{IDENT}_{[\text{low}]}\}$ . Indeed, condition (182) holds for both faithfulness constraints in  $\mathcal{F}$  when the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  in the additional term in (182) is defined as in (iic).

$$(ii) \quad \begin{array}{lll} \text{a.} & \mathbf{a} = \text{t e t d} & \text{b.} & \mathbf{b} = \text{t a d} & \text{c.} & \mathbf{a} = \text{t e t d} \\ & \begin{array}{c} ||| \diagdown \\ \mathbf{d} = \text{t a t} \end{array} & & \begin{array}{c} ||| \\ \mathbf{d} = \text{t a t} \end{array} & & \begin{array}{c} ||| \diagdown \\ \mathbf{b} = \text{t a d} \end{array} \end{array}$$

similar underlying form  $\mathbf{a}$  in correspondence with that same candidate  $\mathbf{c}$  through some relation  $\rho_{\mathbf{a},\mathbf{c}}$ , as represented in (190).



By Definition 6 of the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$ , the assumption  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  ensures the existence of a correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}$  between the two strings  $\mathbf{a}$  and  $\mathbf{b}$ . It is thus natural to assume that  $\rho_{\mathbf{a},\mathbf{c}}$  is the composition  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$  of  $\rho_{\mathbf{a},\mathbf{b}}$  and  $\rho_{\mathbf{b},\mathbf{c}}$ . The existence of this composition candidate is guaranteed by the transitivity axiom (66). The FODC<sup>OT</sup> (170) can thus be specialized as in (191), which will be referred to as the FODC<sub>comp</sub><sup>OT</sup>.

$$(191) \quad \begin{array}{ll} \text{a. If:} & F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \overbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}^{\rho_{\mathbf{a},\mathbf{c}}}) \\ \text{Then:} & F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \\ \text{b. If:} & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \\ \text{Then:} & F(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}_{\rho_{\mathbf{a},\mathbf{c}}}) < F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \end{array}$$

The reasoning in this section is analogous to the reasoning in section 3.2.4, which replaced the general sufficient condition for idempotency provided by the FIC<sup>OT</sup> with the special condition FIC<sub>comp</sub><sup>OT</sup> in terms of the composition candidate.

### 5.5. OT output-drivenness and the triangle inequality

The condition for output-drivenness provided by the FODC<sub>comp</sub><sup>OT</sup> (191) looks like a technical condition without an intuitive interpretation. This section will derive an intuitive metric interpretation of this condition by investigating its relationship with the triangle inequality.

#### 5.5.1. The FTI<sub>comp</sub> entails the FODC<sub>comp</sub><sup>OT</sup> in the general case

Consider two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  which satisfy the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  relative to the similarity order provided by Definition 6. This means that there exists a candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  such that each faithfulness constraint  $F$  in the faithfulness constraint set  $\mathcal{F}$  satisfies the inequality (182), repeated in (192).

$$(192) \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$$

Assume now that the faithfulness constraint  $F$  measures the phonological distance between underlying and surface forms in compliance with the metric triangle inequality, which has been formalized in Section 4.1 as the FTI<sub>comp</sub>. The latter boils down to the inequality (193) in the specific case of the two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  and their composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ .

$$(193) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$$

The inequality (192) allows  $F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  in (193) to be upper bounded with  $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) - F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , yielding the inequality (194).

$$(194) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) - F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) - F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$$

Assume that the antecedent of the first FODC<sub>comp</sub><sup>OT</sup> implication (191a) holds, namely that  $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ . The inequality (194) then entails that  $F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , thus ensuring that the consequent of the implication holds as well. In other words, the inequality (194) entails the first FODC<sub>comp</sub><sup>OT</sup> implication (191a). An analogous reasoning applies to the second FODC<sub>comp</sub><sup>OT</sup> implication

(191b). In conclusion, the  $\text{FTI}_{\text{comp}}$  entails Tesar's  $\text{FODC}_{\text{comp}}^{\text{OT}}$  relative to the similarity order provided by Definition 6, as summarized in the following Proposition 21, which is the main technical result of this chapter. No assumptions are made on the nature of the faithfulness constraints (for instance, they are not required to be categorical) or on the correspondence relations in the candidate set (for instance, they are not required to be one-to-one).

**Proposition 21.** *If a faithfulness constraint  $F$  satisfies the  $\text{FTI}_{\text{comp}}$  and it belongs to the faithfulness constraint set  $\mathcal{F}$  used to define the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  according to Definition 6, then  $F$  satisfies the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  (191) relative to that similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$ .  $\square$*

### 5.5.2. The $\text{FTI}_{\text{comp}}$ is stronger than the $\text{FODC}_{\text{comp}}^{\text{OT}}$ in the general case

Proposition 21 says that the  $\text{FTI}_{\text{comp}}$  entails the  $\text{FODC}_{\text{comp}}^{\text{OT}}$ . The following counterexample shows that the reverse entailment fails in the general case, so that the  $\text{FTI}_{\text{comp}}$  is stronger than the  $\text{FODC}_{\text{comp}}^{\text{OT}}$ . Let  $\ell(\mathbf{a})$  be the length of the string  $\mathbf{a}$ . Assume that the candidate set displays no epenthesis and thus consists of candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  whose surface form  $\mathbf{b}$  is not longer than the underlying form  $\mathbf{a}$ :  $\ell(\mathbf{b}) \leq \ell(\mathbf{a})$ . Since correspondence relations play no role in the counterexample, I omit them and represent candidates simply as pairs of strings. Consider the faithfulness constraint (195) which assigns to a candidate  $(\mathbf{a}, \mathbf{b})$  a number of violations equal to the squared difference of the length of its two strings. This is a proper faithfulness constraint, in the sense that it assigns zero violations to any identity candidate, in compliance with the definitional faithfulness condition (56).

$$(195) \quad F(\mathbf{a}, \mathbf{b}) = (\ell(\mathbf{a}) - \ell(\mathbf{b}))^2$$

This faithfulness constraint  $F$  satisfies both  $\text{FODC}_{\text{comp}}^{\text{OT}}$  implications (191), as their antecedent and consequent are equivalent.<sup>5</sup> Yet,  $F$  fails at the  $\text{FTI}_{\text{comp}}$  for any two candidates  $(\mathbf{a}, \mathbf{b})$  and  $(\mathbf{b}, \mathbf{c})$  and their composition candidate  $(\mathbf{a}, \mathbf{c})$ .<sup>6</sup> This constraint  $F$  thus shows that the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  is weaker than the  $\text{FTI}_{\text{comp}}$ .

### 5.5.3. The $\text{FTI}_{\text{comp}}$ , $\text{FODC}_{\text{comp}}^{\text{OT}}$ , and $\text{FIC}_{\text{comp}}^{\text{OT}}$ are equivalent for categorical constraints

Subsections 5.2.3 and 5.5.1 have shown that the  $\text{FTI}_{\text{comp}}$  entails the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  which in turn entails the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ . Assume now that all correspondence relations in the candidate set are one-to-one, as otherwise output-drivenness fails (as shown in Subsection 5.2.1) and the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  must therefore fail as well. Assume furthermore that each faithfulness constraint in  $\mathcal{F}$  satisfies McCarthy's (strengthened) categoricity conjecture, formalized in section 4.3. Under these assumptions, the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  entails the  $\text{FTI}_{\text{comp}}$ , as shown in Subsection 4.4.3. I conclude that the  $\text{FTI}_{\text{comp}}$ , the  $\text{FODC}_{\text{comp}}^{\text{OT}}$ , and the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  are equivalent, as summarized in the following proposition.

**Proposition 22.** *Assume that the correspondence relations in the candidate set are all one-to-one. Consider a faithfulness constraint  $F$  which is C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone. Assume that  $F$  belongs to the faithfulness subset  $\mathcal{F}$  used to define the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  according to Definition 6. The constraint  $F$  satisfies the  $\text{FTI}_{\text{comp}}$  if and only if it satisfies the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  if and only if it satisfies the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ .  $\square$*

Proposition 22 provides an intuitive interpretation of the rather technical sufficient condition for OT output-drivenness provided by the  $\text{FODC}_{\text{comp}}^{\text{OT}}$ . In fact, it says that the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  simply requires (categorical and monotone) faithfulness constraints to measure the phonological distance between underlying

<sup>5</sup> Consider the first  $\text{FODC}_{\text{comp}}^{\text{OT}}$  implication (191a). Its antecedent and consequent are shown to be equivalent in (i), using the fact that  $x < y$  iff  $x^2 < y^2$ , for any  $x, y \geq 0$

$$\begin{aligned} \text{(i)} \quad F(\mathbf{a}, \mathbf{d}) < F(\mathbf{a}, \mathbf{c}) &\iff (\ell(\mathbf{a}) - \ell(\mathbf{d}))^2 < (\ell(\mathbf{a}) - \ell(\mathbf{c}))^2 \\ &\iff (\ell(\mathbf{a}) - \ell(\mathbf{d})) < (\ell(\mathbf{a}) - \ell(\mathbf{c})) \\ &\iff -\ell(\mathbf{d}) < -\ell(\mathbf{c}) \\ &\iff (\ell(\mathbf{b}) - \ell(\mathbf{d})) < (\ell(\mathbf{b}) - \ell(\mathbf{c})) \\ &\iff (\ell(\mathbf{b}) - \ell(\mathbf{d}))^2 < (\ell(\mathbf{b}) - \ell(\mathbf{c}))^2 \\ &\iff F(\mathbf{b}, \mathbf{d}) < F(\mathbf{b}, \mathbf{c}) \end{aligned}$$

An analogous reasoning holds for the second  $\text{FODC}_{\text{comp}}^{\text{OT}}$  implication (191b).

<sup>6</sup> In fact  $F(\mathbf{a}, \mathbf{c}) = (\ell(\mathbf{a}) - \ell(\mathbf{c}))^2 = [(\ell(\mathbf{a}) - \ell(\mathbf{b})) + (\ell(\mathbf{b}) - \ell(\mathbf{c}))]^2 \geq (\ell(\mathbf{a}) - \ell(\mathbf{b}))^2 + (\ell(\mathbf{b}) - \ell(\mathbf{c}))^2 = F(\mathbf{a}, \mathbf{b}) + F(\mathbf{b}, \mathbf{c})$ , contradicting the  $\text{FTI}_{\text{comp}}$ .

and surface forms in compliance with the metrical axiom of the triangle inequality, as formalized through the  $\text{FTI}_{\text{comp}}$ . Furthermore, Proposition 22 provides a straightforward characterization of the faithfulness constraints which satisfy the  $\text{FODC}_{\text{comp}}^{\text{OT}}$ . In fact, the faithfulness constraints listed in chapter 3 as satisfying the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  are all categorical and monotone. The equivalence between the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  established by Proposition 22 thus ensures that they also all satisfy the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  (under the assumption that all correspondence relations are one-to-one).

#### 5.5.4. The restriction to faithfulness constraints which belong to the subset $\mathcal{F}$

Consider again the two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) = (/p/, [\beta])$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}}) = (/b/, [\beta])$  in (166) (correspondence relations play no role because I am considering singleton segments). As discussed in Subsection 5.1.3, any plausible measure of internal similarity should yield that candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  has less internal similarity than candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ , because the former candidate displays disparities for both voicing and continuancy while the latter displays a disparity for continuancy but not for voicing. Consider the faithfulness constraint  $F = \text{IDENT}_{[\text{voice}]} \vee \text{IDENT}_{[\text{cont}]}$  which is the disjunction of the two identity faithfulness constraints  $\text{IDENT}_{[\text{voice}]}$  and  $\text{IDENT}_{[\text{cont}]}$  for voicing and continuancy. Subsection 5.2.4 has shown that  $F$  fails at the  $\text{FODC}_{\text{comp}}^{\text{OT}}$ . Yet, Proposition 11 in Subsection 3.5.6 states that  $F$  succeeds at the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ . Furthermore,  $F$  is obviously C-categorical, as it is the disjunction of two identity constraints which are both C-categorical. The fact that  $F$  satisfies the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  but not the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  is *not* a counterexample to the entailment from the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  to the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  ensured by Proposition 22. In fact, that proposition only looks at the faithfulness constraints which crucially belong to the faithfulness constraint set  $\mathcal{F}$  used to define the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  according to Definition 6. Crucially, the disjunctive faithfulness constraint  $F = \text{IDENT}_{[\text{voice}]} \vee \text{IDENT}_{[\text{cont}]}$  cannot belong to the set  $\mathcal{F}$ . In fact, the inequality (196) required by Definition 6 for the similarity inequality  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  to hold fails for this specific faithfulness constraint  $F$ .<sup>7</sup> In other words, if  $F$  did belong to  $\mathcal{F}$ , the two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  would not be comparable relative to the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$ , contrary to intuitions.

$$(196) \quad \underbrace{F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})}_{=F(/p/, [\beta])=1} \not\leq \underbrace{F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})}_{=F(/b/, [\beta])=1} + \underbrace{F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})}_{=F(/p/, [b])=1}$$

The reason why Definition 6 of the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  has been parameterized by a faithfulness constraint subset  $\mathcal{F}$  possibly smaller than the entire faithfulness constraint set is precisely to allow for the possibility that certain faithfulness constraints (and in particular those derived from other more basic faithfulness constraints through operations such as local disjunction) not be considered in the computation of similarity.

#### 5.5.5. Summary

This section has reconstructed Tesar's (2013) theory of output-drivenness. The sufficient condition for output-drivenness provided by his  $\text{FODC}^{\text{OT}}$  has been shown to admit a metric interpretation: it effectively requires the faithfulness constraints (which belong to the faithfulness constraint set  $\mathcal{F}$  used to compare similarity) to measure the phonological distance between underlying and surface forms in compliance with the metric triangle inequality. This interpretation holds under McCarthy's (strengthened) categoricity conjecture. Because of this reinterpretation, the  $\text{FODC}^{\text{OT}}$  turns out to be equivalent to the sufficient condition for idempotency provided by the  $\text{FIC}^{\text{OT}}$  (for faithfulness constraints which are categorical and belong to  $\mathcal{F}$ ).

### 5.6. Output-drivenness in Harmonic Grammar

This section extends the theory of output-drivenness from OT to HG. This extension will provide a more pristine view of the relationship between output-drivenness and the triangle inequality, which does not rely on categoricity.

<sup>7</sup> I am assuming that the correspondence relation  $\rho_{\mathbf{a}, \mathbf{b}}$  in the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}}) = (/p/, [b])$  does put the singleton underlying and surface segments in correspondence. If that is not the case, then the inequality (196) would indeed succeed for  $F = \text{IDENT}_{[\text{voice}]} \vee \text{IDENT}_{[\text{cont}]}$  but it would fail for  $F = \text{MAX}$  and  $F = \text{DEP}$ .

### 5.6.1. A condition on the faithfulness constraint set: the $\text{FODC}_{\text{comp}}^{\text{HG}}$

Subsection 5.2.2 has recalled Tesar’s (2013) Proposition 18, which provides guarantees for the output-drivenness of the grammars in an OT typology under the assumption that each faithfulness constraint satisfies the  $\text{FODC}^{\text{OT}}$  (170). The following Proposition 23 provides the HG analogue of Tesar’s OT Proposition 18. The proof is analogous to that of Proposition 13 for HG idempotency and it is provided in Appendix A.3. The implication (197) will be referred to as the *HG faithfulness output-drivenness condition* ( $\text{FODC}^{\text{HG}}$ ). Proposition 23 makes no assumptions on the markedness constraints, on the nature of the faithfulness constraints (for instance, it does not require them to be categorical), on the correspondence relations in the candidate set (for instance, it does not require them to be one-to-one), or on the similarity order  $\leq_{\text{sim}}$  (which is indeed left completely unspecified).

**Proposition 23.** *Assume that, for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , for every other candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , the candidate set also contains a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  different from  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  such that the following condition (197) holds for any faithfulness constraint  $F$  in the constraint set.*

(197) *For every choice of the constant  $\xi$  (with no restrictions on its sign):*

$$\text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi$$

$$\text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi$$

*Then, the HG grammar corresponding to any weighting of the constraint set is output-driven relative to the similarity order  $\leq_{\text{sim}}$ .  $\square$*

Let’s now consider the special case where the similarity order is defined as in Subsection 5.4.2. The similarity condition  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  thus means that there exists some candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  which puts the two underlying forms  $\mathbf{a}$  and  $\mathbf{b}$  in correspondence and validates the crucial inequality (182) for every faithfulness constraint  $F$  in the faithfulness constraint subset  $\mathcal{F}$ . By reasoning as in Subsection 5.4.5, the  $\text{FODC}^{\text{HG}}$  (197) can be specialized to the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  (198), by assuming that the candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  which appears on the left-hand side of the inequality in the consequent is the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  of the two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , whose existence is guaranteed by the transitivity axiom (66).

(198) *For every choice of the constant  $\xi$  (with no restrictions on its sign):*

$$\text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi$$

$$\text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi$$

The sufficient condition for output-drivenness provided by the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  is *tight*: any faithfulness constraint which fails the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  admits an elementary counterexample where output-drivenness fails.

### 5.6.2. The $\text{FODC}^{\text{HG}}$ entails the $\text{FIC}^{\text{HG}}$ in the general case

Subsection 5.2.3 has shown that the sufficient condition for OT output-drivenness provided by the  $\text{FODC}^{\text{OT}}$  entails the sufficient condition for OT idempotency provided by the  $\text{FIC}^{\text{OT}}$ , because the latter coincides with the former in the special case where  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  is the identity candidate. An analogous reasoning holds in the case of HG. The sufficient condition for HG output-drivenness provided by the  $\text{FODC}^{\text{HG}}$  (197) reduces to (199) when  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  is the identity candidate. The latter is the sufficient condition for HG idempotency provided by the  $\text{FIC}^{\text{HG}}$  (112), as derived in Section 4.2.3.<sup>8</sup>

(199) *For every choice of the constant  $\xi$  (with no restrictions on its sign):*

$$\text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi$$

$$\text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi$$

In conclusion, the  $\text{FODC}^{\text{HG}}$  entails the  $\text{FIC}^{\text{HG}}$  in the general case, no matter how the similarity order in the definition of output-drivenness is defined. This conclusion matches the fact that output-drivenness entails idempotency independently of the grammatical framework, as noted in Subsection 5.1.2.

<sup>8</sup> As explained in footnote 2, it makes no difference whether the constant  $\xi$  in (112)/(199) is restricted to be non-negative or allowed to be negative. That is of course not the case for (197): the fact that  $\xi$  is allowed to be negative makes it a stronger condition.

### 5.6.3. The $FODC^{HG}$ entails the $FODC^{OT}$ in the general case

Subsection 5.2.2 has introduced Tesar’s sufficient condition  $FODC^{OT}$  for OT output-drivenness, repeated in (200) for ease of comparison.

$$(200) \quad \begin{array}{ll} \text{a. If:} & F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \\ & \text{Then: } F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \\ \\ \text{b. If:} & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \\ & \text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) < F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \end{array}$$

These two implications (200a) and (200b) are closely related: if we were to replace “<” with “≤”, then one would be the counter-positive of the other and we could retain only one of the two implications. This intuition is brought out by the following Proposition 24: the two implications (200a) and (200b) are jointly equivalent to the condition (201). The latter indeed coincides with the second implication (200b) where “<” has been replaced with “≤” at the price of adding a small constant  $\xi$  at the right-hand side. The proof of this equivalence is straightforward but tedious, and it is therefore relegated to Appendix A.4.

**Proposition 24.** *The two  $FODC^{OT}$  implications (200) are jointly equivalent to the condition:*

$$(201) \quad \begin{array}{l} \text{For every choice of the constant } \xi \text{ between } -1 \text{ and } +1 \text{ (both excluded):} \\ \text{If:} \quad F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi \\ \text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi \end{array}$$

for any faithfulness constraint  $F$ . □

The restatement (201) of the  $FODC^{OT}$  shows that it is actually entailed by the sufficient condition  $FODC^{HG}$  (197) for HG output-drivenness in the general case, matching the fact that HG typologies can be larger than OT typologies.

### 5.6.4. The $FODC_{comp}^{HG}$ , $FIC^{HG}$ , and $F_{TI}_{comp}$ are equivalent in the general case

The  $FODC_{comp}^{HG}$  entails the  $FIC_{comp}^{HG}$  in the general case (as seen in Subsection 5.6.2). The  $FIC_{comp}^{HG}$  is in turn equivalent to the triangle inequality formalized through the  $F_{TI}_{comp}$  (as seen in Subsection 4.2.5). Thus, the  $FODC_{comp}^{HG}$  entails the  $F_{TI}_{comp}$ . This entailment holds under no assumptions on the faithfulness constraints. To investigate the reverse entailment, consider a faithfulness constraint  $F$  which belongs to the faithfulness constraint set  $\mathcal{F}$  used to define the similarity order  $\leq_{sim}^{\mathcal{F}}$  according to Definition 6 in Subsection 5.4.2. Consider two candidates such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ . As seen in Subsection 5.5.1, the  $F_{TI}_{comp}$  entails the inequality (194), repeated in (202). And the latter in turn straightforwardly entails the  $FODC_{comp}^{HG}$ .

$$(202) \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}, \rho_{\mathbf{b},\mathbf{c}}) - F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) - F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$$

In conclusion, the  $FODC_{comp}^{HG}$  relative to the similarity order  $\leq_{sim}^{\mathcal{F}}$  provided by the Definition 6 is equivalent to both the  $FIC_{comp}^{HG}$  and the  $F_{TI}_{comp}$ , as summarized in the following Proposition 21. This equivalence holds for any faithfulness constraint which belongs to  $\mathcal{F}$ . No additional assumptions are made neither on the nature of the faithfulness constraints (for instance, they are not required to be categorical) nor on the correspondence relations in the candidate set (for instance, they are not required to be one-to-one).

**Proposition 25.** *The  $FODC_{comp}^{HG}$  relative to the similarity order  $\leq_{sim}^{\mathcal{F}}$  is equivalent to both the  $FIC_{comp}^{HG}$  and the  $F_{TI}_{comp}$  for any faithfulness constraint which belongs to the faithfulness set  $\mathcal{F}$  used to define the similarity order.* □

This proposition provides an intuitive interpretation of the rather technical sufficient condition for HG output-drivenness provided by the  $FODC_{comp}^{HG}$ . In fact, it says that the  $FODC_{comp}^{HG}$  simply requires the faithfulness constraints (which belong to  $\mathcal{F}$ ) to measure the phonological distance between underlying and surface forms in compliance with the metrical axiom of the triangle inequality, as formalized through the  $F_{TI}_{comp}$ .



### 5.6.5. The $\text{FODC}_{\text{comp}}^{\text{HG}}$ , $\text{FODC}_{\text{comp}}^{\text{OT}}$ and $\text{FIC}_{\text{comp}}^{\text{OT}}$ are equivalent for categorical constraints

The  $\text{FODC}_{\text{comp}}^{\text{HG}}$  entails the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  in the general case (as seen in Subsection 5.6.3). The  $\text{FODC}_{\text{comp}}^{\text{OT}}$  in turn entails the sufficient condition for OT idempotency provided by the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  (as seen in Subsection 5.2.3). Thus, the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  entails the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ . This entailment holds under no assumptions on the faithfulness constraints. To investigate the reverse entailment, assume that all correspondence relations in the candidate set are one-to-one, as otherwise output-drivenness fails (as seen in Subsection 5.2.1) and the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  thus fails as well. Consider a faithfulness constraint which satisfies McCarthy's (strengthened) categoricity conjecture, formalized in section 4.3. Under these assumptions, the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  entails the  $\text{FTI}_{\text{comp}}$  (as seen in Subsection 4.4.3). The  $\text{FTI}_{\text{comp}}$  in turn entails the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for every faithfulness constraint which belongs to the faithfulness constraint set  $\mathcal{F}$  used to define the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  (as seen in Subsection 5.6.4). We have thus obtained the following equivalence among the  $\text{FODC}_{\text{comp}}^{\text{HG}}$ , the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  and the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ .

**Proposition 26.** *Assume that the correspondence relations in the candidate set are all one-to-one. Consider a faithfulness constraint  $F$  which is  $C$ -categorical; or  $I$ -categorical and  $O$ -monotone; or  $O$ -categorical and  $I$ -monotone. Assume that  $F$  belongs to the faithfulness set  $\mathcal{F}$  used to define the similarity order  $\leq_{\text{sim}}^{\mathcal{F}}$  according to Definition 6. The constraint  $F$  satisfies the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  if and only if it satisfies the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  if and only if it satisfies the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ .  $\square$*

This Proposition 22 provides a straightforward characterization of the faithfulness constraints which satisfy the  $\text{FODC}_{\text{comp}}^{\text{HG}}$ . In fact, the faithfulness constraints listed in chapter 3 as satisfying the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  are all categorical and monotone. The equivalence between the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  established by Proposition 26 thus ensures that they also all satisfy the  $\text{FODC}_{\text{comp}}^{\text{HG}}$  relative to  $\leq_{\text{sim}}^{\mathcal{F}}$  (under the assumption that all correspondence relations are one-to-one and that they belong to the subset  $\mathcal{F}$ ).

### 5.6.6. Summary

This section has completed the theory of output-drivenness within the OT and HG implementations of constraint-based phonology. Output-drivenness follows from conditions on the faithfulness constraints, namely the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  and the  $\text{FODC}_{\text{comp}}^{\text{HG}}$ . These conditions are equivalent for faithfulness constraints which satisfy McCarthy's (strengthened) categoricity conjecture, because both conditions can be interpreted as requiring faithfulness constraints to measure phonological distances in compliance with the triangle inequality, formalized as the  $\text{FTI}_{\text{comp}}$ .

## A. Proofs

### A.1. Proof of proposition 18

**PROPOSITION 18** *Assume that, for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ , for any other candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ , the candidate set also contains a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  such that the two following implications (170) hold for every faithfulness constraint  $F$  in the constraint set.*

$$(203) \quad \begin{array}{ll} a. & \text{If: } F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) \\ & \text{Then: } F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) \\ b. & \text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}}) \\ & \text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) < F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \end{array}$$

*Then, the OT grammar corresponding to any ranking of the constraint set is output-driven relative to the similarity order  $\leq_{\text{sim}}$ , no matter what the markedness constraints look like.  $\square$*

**PROOF.** Suppose that the OT grammar corresponding to a certain ranking maps the underlying form  $\mathbf{a}$  to the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$ . This means that the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  beats any other candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  for any other surface form  $\mathbf{c}$  according to that ranking. This means in turn that one constraint which correctly prefers the winner candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  over the loser candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  is ranked above every constraint which instead prefers  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  over  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  as in (204).

$$(204) \quad \begin{array}{l} C \\ | \\ C_1, C_2, \dots \end{array} \quad \begin{array}{l} \text{a constraint which prefers } (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \text{ over } (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) \\ \text{all the constraints which prefer } (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) \text{ over } (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \end{array}$$



(207). The inequality (213) trivially holds (as an identity), as faithfulness constraints assign zero violations to the identity candidate  $(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a},\mathbf{a}})$ . The latter inequality coincides with the definitional inequality (209) in the case  $\mathbf{a} = \mathbf{b}$  and thus ensures that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ , as required by reflexivity.

$$(213) \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + F(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a},\mathbf{a}})$$

Let me show that  $\leq_{\text{sim}}^{\mathcal{F}}$  is *antisymmetric*, namely that it satisfies the implication (214) for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ .

$$(214) \quad \begin{array}{l} \mathbf{If:} \quad (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \text{ and } (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \\ \mathbf{Then:} \quad (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) = (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}). \end{array}$$

By the definition of  $\leq_{\text{sim}}^{\mathcal{F}}$ , the antecedent of the implication (214) means that there exist two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{a}, \rho_{\mathbf{b},\mathbf{a}})$  which validate the inequalities (215).

$$(215) \quad \begin{array}{l} \text{a. } F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \\ \text{b. } F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \geq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + F(\mathbf{b}, \mathbf{a}, \rho_{\mathbf{b},\mathbf{a}}) \end{array}$$

The chain of inequalities in (216) then holds. In step (216a), I have used the inequality (215a) together with the fact that constraint violations are non-negative. In step (216b), I have used the inequality (215b).

$$(216) \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \stackrel{(a)}{\geq} F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \stackrel{(b)}{\geq} F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + F(\mathbf{b}, \mathbf{a}, \rho_{\mathbf{b},\mathbf{a}})$$

The inequality (216) thus derived, together with the non-negativity of constraint violations, implies that  $F(\mathbf{b}, \mathbf{a}, \rho_{\mathbf{b},\mathbf{a}}) = 0$ . An analogous reasoning shows that  $F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) = 0$ . The two inequalities (215) thus become  $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  and  $F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \geq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ , which entail the identity (217). Since the latter identity holds for every faithfulness constraint  $F$  in  $\mathcal{F}$  and since the faithfulness constraint subset  $\mathcal{F}$  is by hypothesis complete relative to the candidate set, I conclude that the two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  coincide, as required by the consequent of the implication (214).

$$(217) \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) = F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$$

Finally, let me show that  $\leq_{\text{sim}}^{\mathcal{F}}$  is *transitive*, namely that it satisfies the implication (218) for any three candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ ,  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , and  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}})$

$$(218) \quad \begin{array}{l} \mathbf{If:} \quad (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \text{ and } (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}}) \\ \mathbf{Then:} \quad (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}}) \end{array}$$

By the definition of  $\leq_{\text{sim}}^{\mathcal{F}}$ , the antecedent of the implication (218) means that there exist two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which validate the inequalities (219).

$$(219) \quad \begin{array}{l} \text{a. } F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \\ \text{b. } F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \geq F(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \end{array}$$

Consider the candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  which is the composition of the two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , whose existence is guaranteed by the transitivity axiom (208). The chain of inequalities (220) then holds. In step (220a), I have used the inequality (219a). In step (220b), I have used the inequality (219b). Finally in step (220c), I have used the FTI<sub>comp</sub> (210).

$$(220) \quad \begin{array}{l} F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \stackrel{(a)}{\geq} F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \\ \stackrel{(b)}{\geq} F(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}}) + F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) + F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) \\ \stackrel{(c)}{\geq} F(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}}) + F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \end{array}$$

The inequality obtained in (220) ensures that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}})$ , as required by the consequent of the implication (218).  $\square$

### A.3. Proof of proposition 23

**PROPOSITION 23** Assume that, for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , for every candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , the candidate set also contains a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  different from  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  such that the  $FODC^{HG}$  repeated in (221) holds for any faithfulness constraint  $F$ .

(221) For every choice of the constant  $\xi$  (with no restrictions on its sign):

$$\begin{aligned} \text{If:} \quad & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi \\ \text{Then:} \quad & F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi \end{aligned}$$

Then, the HG grammar corresponding to any weighting of the constraint set is output-driven relative to the similarity order  $\leq_{sim}$ .  $\square$

PROOF. The proof is similar to the proof of Proposition 13 in Appendix A.1. Suppose that the HG grammar  $G_\theta$  corresponding to some weighting  $\theta$  fails at the output-drivenness implication (162) for two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , as stated in (222): the grammar  $G_\theta$  maps the underlying form  $\mathbf{a}$  to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  with less internal similarity, as required by the antecedent of (162); but it fails to map the underlying form  $\mathbf{b}$  to the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  with more internal similarity, as required by the consequent of (162).

- (222)  $G_\theta$  fails at output-drivenness on candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  iff there exists a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that:
- $G_\theta(\mathbf{a}) = (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ ;
  - $G_\theta(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \neq (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ .

By assumption, the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  comes with a companion candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  different from  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ . The “if-and-only-if” statement (222) can thus be weakened into the “if” statement (223).

- (223) If  $G_\theta$  fails at output-drivenness on candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , there exists a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that:
- $\theta$  prefers  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ ,
  - $\theta$  prefers  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  to  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ .

Condition (223) can be made explicit as in (224) in terms of the numbers of constraint violations. The sums run over a generic markedness constraint  $M$  with weight  $\theta_M$  and a generic faithfulness constraint  $F$  with weight  $\theta_F$ .

- (224) If  $G_\theta$  fails at output-drivenness on candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , there exists a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that:
- $$\begin{aligned} \sum_M \theta_M M(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \sum_F \theta_F F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) &< \\ &< \sum_M \theta_M M(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \end{aligned}$$
  - $$\begin{aligned} \sum_M \theta_M M(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) &< \\ &< \sum_M \theta_M M(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \sum_F \theta_F F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \end{aligned}$$

Taking advantage of the fact that markedness constraints are blind to underlying forms by (57), condition (224) can be rewritten as in (225) with the position  $\xi = \sum_M \theta_M M(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) - \sum_M \theta_M M(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) = \sum_M \theta_M M(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) - \sum_M \theta_M M(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ .

- (225) If  $G_\theta$  fails at output-drivenness on candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ , there exists a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that:
- $$\sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) > \sum_F \theta_F F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi$$
  - $$\sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \sum_F \theta_F F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi$$

In conclusion, output-drivenness holds for the HG grammar corresponding to any constraint weighting provided the two conditions (225a) and (225b) can never be satisfied both, no matter the choice of the weights  $\theta_F$  and the constant  $\xi$ . In other words, it suffices to assume that for every candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  different from  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  there exists a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  different from  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  such that:

(226) For every choice of the constant  $\xi$ , for every choice of the weights  $\theta_F$ :

$$\begin{aligned} \text{If:} \quad & \sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \sum_F \theta_F F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi \\ \text{Then:} \quad & \sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq \sum_F \theta_F F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi \end{aligned}$$

To conclude the proof, condition (226) can be shown to be equivalent to the FODC<sup>HG</sup> (221) by reasoning as at the end of the proof of Proposition 13 to show that condition (148) is equivalent to the FIC<sup>HG</sup> (142).  $\square$

#### A.4. Proof of proposition 24

**PROPOSITION 24** *The two FODC<sup>OT</sup> implications repeated in (227)*

$$\begin{aligned} (227) \quad a. \quad & \text{If:} \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \\ & \text{Then:} \quad F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \\ b. \quad & \text{If:} \quad F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \\ & \text{Then:} \quad F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) < F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \end{aligned}$$

are jointly equivalent to the following condition:

(228) *For every choice of the constant  $\xi$  between  $-1$  and  $+1$  (both excluded):*

$$\begin{aligned} \text{If:} \quad & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi \\ \text{Then:} \quad & F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi \end{aligned}$$

for any faithfulness constraint  $F$ .  $\square$

**PROOF.** Let me show that the two FODC<sup>OT</sup> implications (227) jointly entail the implication (228). Thus, assume that the antecedent of the latter implication holds for some  $\xi$ . I distinguish two cases, depending on whether  $\xi$  is (strictly) smaller than 0 or not. Let me start with the former case, stated in (229a).

$$\begin{aligned} (229) \quad a. \quad & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi \text{ with } -1 < \xi < 0 \\ b. \quad & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \\ c. \quad & F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}\rho_{\mathbf{b},\mathbf{c}}}) < F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \\ d. \quad & F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}\rho_{\mathbf{b},\mathbf{c}}}) \leq F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi \end{aligned}$$

Since  $\xi$  is strictly negative, (229a) entails the strict inequality (229b). The latter in turn coincides with the antecedent of the second FODC<sup>OT</sup> implication (227b), which therefore ensures that its consequent holds as well, repeated in (229c). Since  $\xi$  is larger than  $-1$  and constraint violations are integers, (229c) in turn entails (229d), which is the desired consequent of the implication (228). Note that this reasoning has used only the second FODC<sup>OT</sup> implication (227b).

Consider next the complementary case where the antecedent of the implication (228) holds with a nonnegative  $\xi$ , as stated in (230a).

$$\begin{aligned} (230) \quad a. \quad & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi \text{ with } 0 \leq \xi < +1 \\ b. \quad & F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \\ c. \quad & F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \\ d. \quad & F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi \geq F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \end{aligned}$$

Since  $\xi$  is smaller than  $+1$  and constraint violations are integers, (229a) entails (229b). The latter in turns says that the consequent of the first FODC<sup>OT</sup> implication (227a) fails. The antecedent must therefore fail as well, as stated in (229c). Since  $\xi$  is nonnegative, the latter in turn entails (229d), which is the desired consequent of the implication (228). Note that this reasoning has used only the first FODC<sup>OT</sup> implication (227a).

Next, let me show that condition (228) with  $0 < \xi < +1$  in turn entails the first  $\text{FODC}^{\text{OT}}$  implication (227a). In fact, suppose that the antecedent of the latter implication holds, namely that  $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$ . Since  $\xi$  is smaller than  $+1$  and constraint violations are integers, the latter entails that  $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) + \xi < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}}, \rho_{\mathbf{b}, \mathbf{c}})$ . The consequent of the implication (228) thus fails. Its antecedent must therefore fail as well, namely  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) \geq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}}) + \xi$ . The latter entails that  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) > F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ , establishing the consequent of the first  $\text{FODC}_{\text{comp}}^{\text{OT}}$  implication (227a). An analogous reasoning shows that condition (228) with  $-1 < \xi < 0$  entails the second  $\text{FODC}_{\text{comp}}^{\text{OT}}$  implication (227b).  $\square$



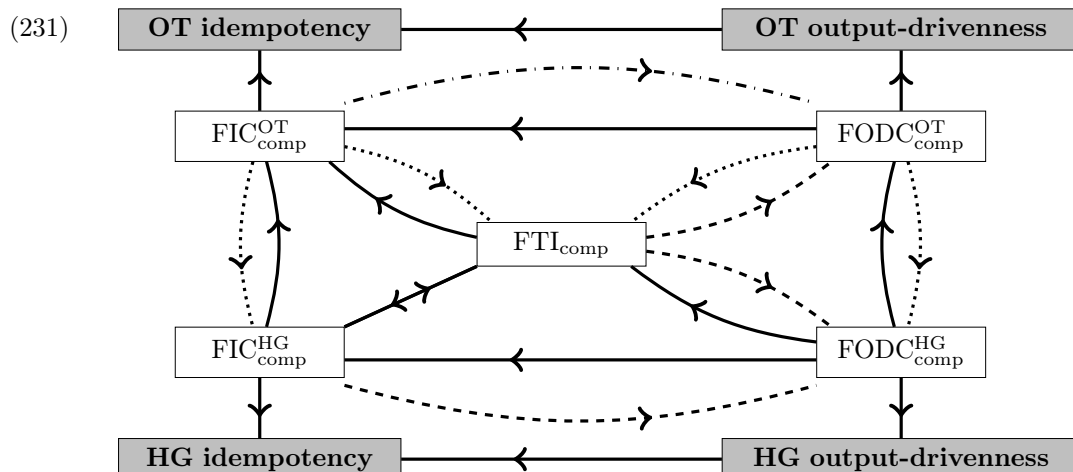
## Implications for phonology and its acquisition

Section 6.1 summarizes the main results obtained in the preceding chapters, highlighting the place of the faithfulness triangle inequality at the center of the theory of idempotency and output-drivenness. The rest of the chapter then speculates on the applications of the theory of idempotency and output-drivenness for phonological theory and for the acquisition of phonology. To start, section 6.2 looks into Tesar’s claim that output-drivenness corresponds to transparent phonology while failure of output-drivenness corresponds to phonological opacity. This connection is brought out by formalizing the relationship (implicit in much literature on phonological opacity) between counter-feeding and chain shifts and between counter-bleeding and saltations. Section 6.3 formulates the conjecture that a dual of the notion of output-drivenness, called here *input-drivenness*, might provide a better match with the distinction between opacity and transparency.

The rest of the chapter zooms in on the implications of the theory of idempotency for chain shifts. From the perspective of phonology, the crucial question raised by chain shifts is *how to get them*. Section 6.4 summarizes various OT approaches to chain shifts, showing that they differ for the choice of a faithfulness constraint which fails at the sufficient faithfulness condition for idempotency provided by the FIC in chapter 3. On the other hand, the type of opacity introduced by chain shifts has been argued to make the learning problem harder, particularly for the initial stage of phonotactic learning. From the perspective of learnability, the crucial question seems therefore to be the opposite, namely *how to avoid chain shifts*. Section 6.5 tries to reconcile this tension between phonology and learnability, by formulating the conjecture that attested chain shifts turn out to be *benign* for learnability.

### 6.1. Summary: idempotency, output-drivenness, and the triangle inequality

The results obtained in the preceding chapters can be summarized through the entailments in (231). The solid arrows correspond to entailments which hold for arbitrary faithfulness constraints. The dotted arrows correspond to entailments which hold for faithfulness constraints which satisfy the categoricity-plus-monotonicity conjecture formulated in section 4.3. The dashed arrows correspond to entailments which hold for faithfulness constraints that belong to the faithfulness constraint set  $\mathcal{F}$  used in the axiomatic definition of phonological similarity along the lines of section 5.4. The dotted/dashed arrows correspond to entailments which hold for faithfulness constraints which belong to  $\mathcal{F}$  and are categorical-plus-monotone.





The diagram nicely summarizes the main conclusion obtained: the triangle inequality for faithfulness constraints formalized in section 4.1 through the  $\text{FTI}_{\text{comp}}^{\text{comp}}$  sits at the heart of the theory of idempotency and output-drivenness. The results obtained, summarized in the diagram (231), have a number of implications summarized in the rest of this section.

### 6.1.1. Implications for the interpretation of faithfulness constraint conditions

The four sufficient conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FIC}_{\text{comp}}^{\text{HG}}$ ,  $\text{FODC}_{\text{comp}}^{\text{OT}}$ , and  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for OT and HG idempotency and output-drivenness are abstract and technical conditions which do not seem *prima facie* to admit an intuitive interpretation. The equivalences with the  $\text{FTI}_{\text{comp}}$  at the center of the scheme (231) is thus useful because it provides an intuitive metric interpretation of these four abstract conditions. According to this interpretation, these conditions simply require the faithfulness constraints to measure the phonological distance between underlying and surface forms in compliance with the triangle inequality which characterizes metrical distances. This metric interpretation holds under no additional assumptions for the HG conditions  $\text{FIC}_{\text{comp}}^{\text{HG}}$  and  $\text{FODC}_{\text{comp}}^{\text{HG}}$  while it requires categoricity-plus-monotonicity for the OT conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and  $\text{FODC}_{\text{comp}}^{\text{OT}}$  (for the case of output-drivenness, it also requires membership to the faithfulness constraint set  $\mathcal{F}$  used to measure phonological similarity).

### 6.1.2. Implications for the relationship between idempotency and output-drivenness

Output-drivenness entails idempotency: if phonological forms are only modified to meet the phonotactics (output-drivenness), phonotactically licit forms are faithfully realized (idempotency). Indeed, the conditions  $\text{FODC}_{\text{comp}}^{\text{OT}}$  and  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for output-drivenness entail the corresponding conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and  $\text{FIC}_{\text{comp}}^{\text{HG}}$  for idempotency in the general case. Consider a faithfulness constraint which belongs to the faithfulness constraint set  $\mathcal{F}$  used to measure phonological similarity in the definition of output-drivenness. For such a faithfulness constraint, the scheme (231) then says that the conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and  $\text{FIC}_{\text{comp}}^{\text{HG}}$  for idempotency are actually equivalent to the corresponding conditions  $\text{FODC}_{\text{comp}}^{\text{OT}}$  and  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for output-drivenness (in the case of OT, the equivalence requires the additional assumption that the faithfulness constraint satisfies the categoricity-plus-monotonicity conjecture). This conclusion says that output-drivenness is stronger than idempotency (in the sense that there exist idempotent grammars which fail at output-drivenness) only if the relation of similarity underlying output-drivenness is blind to some of the faithfulness constraints used to define the typology, namely only if similarity is measured relative to a faithfulness constraint set  $\mathcal{F}$  which is smaller than or different from the faithfulness constraint set used to define the typology.

### 6.1.3. Implications for the relationship between OT and HG

HG typologies are larger than OT typologies. The conditions needed to discipline all grammars in a typology to satisfy idempotency or output-drivenness are thus expected to be stronger in HG than in OT. Indeed, the conditions  $\text{FIC}_{\text{comp}}^{\text{HG}}$  and  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for idempotency and output-drivenness in HG entail the corresponding conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and  $\text{FODC}_{\text{comp}}^{\text{OT}}$  in OT in the general case. Yet, the scheme (231) says that the OT conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$  and  $\text{FODC}_{\text{comp}}^{\text{OT}}$  for idempotency and output-drivenness are actually equivalent to the corresponding HG conditions  $\text{FIC}_{\text{comp}}^{\text{HG}}$  and  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for faithfulness constraints which are categorical-plus-monotone, thus distilling yet another implication of McCarthy's (2003) categoricity conjecture.

### 6.1.4. Implications for the analysis of faithfulness constraints

Which faithfulness constraints satisfy the conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FIC}_{\text{comp}}^{\text{HG}}$ ,  $\text{FODC}_{\text{comp}}^{\text{OT}}$ , or  $\text{FODC}_{\text{comp}}^{\text{HG}}$  for idempotency and output-drivenness in OT and HG? The answer to this question is non-trivial. Indeed, the entire chapter 3 of Tesar's book is devoted to establishing the  $\text{FODC}_{\text{comp}}^{\text{OT}}$  for just the three constraints MAX, DEP, and IDENT. The scheme (231) affords a substantial simplification. Since the four conditions  $\text{FIC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FIC}_{\text{comp}}^{\text{HG}}$ ,  $\text{FODC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FODC}_{\text{comp}}^{\text{HG}}$  are equivalent for categorical-plus-monotone faithfulness constraints, it suffices to focus on the  $\text{FIC}_{\text{comp}}^{\text{OT}}$ , which is arguably the simplest of the four conditions, as revealed by the fact that it is the weakest one in the general case. Sections 3.3-3.5 have investigated this condition for a large variety of faithfulness constraints which have been proposed within McCarthy and Prince's (1995) Correspondence Theory and its more recent developments. Since those constraints all satisfy the categoricity-plus-monotonicity requirement, the scheme (231) says that those results concerning

the faithfulness constraints which satisfy the  $\text{FIC}_{\text{comp}}^{\text{OT}}$  translate straightforwardly into results concerning the faithfulness constraints which satisfy the other three conditions  $\text{FIC}_{\text{comp}}^{\text{HG}}$ ,  $\text{FODC}_{\text{comp}}^{\text{OT}}$ ,  $\text{FODC}_{\text{comp}}^{\text{HG}}$ . A measure of the improvement obtained is provided by the fact that a large array of faithfulness constraints (beyond the three considered by Tesar) are shown in a snap to satisfy the  $\text{FODC}_{\text{comp}}^{\text{OT}}$ .

## 6.2. Output-drivenness and opacity

Tesar’s output-drivenness is meant to capture at the *extensional* level of phonological mappings the distinction between *transparent* and *opaque* patterns which is usually drawn at the *intensional* level of conditions on rule ordering. Ideally, we would like to equate transparency with output-drivenness and opacity with non-output-drivenness. This section elaborates on this claim by making explicit through propositions 27 and 28 the connection (implicit in much of the literature on opacity) between counter-feeding and chain shifts and between counter-bleeding and saltations.

### 6.2.1. Counter-feeding and chain shifts

As in (most of) the rule-based literature on opacity (Baković 2011 and references therein), let me focus for simplicity on derivations consisting of only two ordered rules, which will be denoted by  $\mathbb{A}$  and  $\mathbb{B}$ . Consider a target derivation (232a) which takes an underlying form  $\mathbf{a}$  and applies rule  $\mathbb{A}$  first followed by rule  $\mathbb{B}$ . Consider the derivation (232b) where the order of the two rules  $\mathbb{A}$  and  $\mathbb{B}$  is swapped. Rule  $\mathbb{B}$  is said to *counter-feed* rule  $\mathbb{A}$  relative to the derivation (232a) provided rule  $\mathbb{A}$  applies vacuously in that derivation but it applies non-vacuously in the swapped derivation (232b). In other words,  $\mathbb{B}$  counter-feeds  $\mathbb{A}$  relative to a derivation (232a) if and only if  $\mathbb{B}$  feeds  $\mathbb{A}$  relative to the swapped derivation (232b).

- (232) a. UR:  $\mathbf{a}$   
           rule  $\mathbb{A}$ : VACUOUS  
           rule  $\mathbb{B}$ : ...  
       b. UR:  $\mathbf{a}$   
           rule  $\mathbb{B}$ : ...  
           rule  $\mathbb{A}$ : NON VACUOUS

As an example (taken from Baković 2011), let  $\mathbb{A}$  be a rule which palatalizes consonants before front vowels and let  $\mathbb{B}$  be a rule which deletes the first vowel of a hiatus. Deletion counter-feeds palatalization in the derivation in (233a), as shown by comparison with the swapped derivation in (233b).

- (233) a. UR: /tui/  
           rule  $\mathbb{A}$  = Palatalization: tui (VACUOUS)  
           rule  $\mathbb{B}$  = Deletion: ti  
       b. UR: /tui/  
           rule  $\mathbb{B}$  = Deletion: ti  
           rule  $\mathbb{A}$  = Palatalization: tʃi (NON VACUOUS)

Counter-feeding is said to yield *under-application opacity*: the palatalization rule  $\mathbb{A}$  has under-applied in the derivation (233a), as the surface form [ti] obtained through that derivation would undergo palatalization if it were fed back to the phonology as the underlying form /ti/. This intuition can be restated as follows: the SPE phonology corresponding to the two ordered rules  $\langle \mathbb{A}, \mathbb{B} \rangle = \langle \text{Pal}, \text{Del} \rangle$  yields the chain shift (234) whereby /tui/ is mapped to [ti] but /ti/ is mapped to [tʃi].

- (234) tui  $\longrightarrow$  ti  $\longrightarrow$  tʃi

As stated by the following proposition, this conclusion holds in general under mild assumptions: counter-feeding opacity entails chain shifts. The “mild assumptions” mentioned in the informal statement of the proposition are made explicit in appendix A.1.

**Proposition 27.** *Consider a derivation which starts at an underlying form  $\mathbf{a}$  and applies rule  $\mathbb{A}$  first followed by rule  $\mathbb{B}$ , yielding a surface form  $\mathbf{b}$ . If it is a counter-feeding derivation, then the derivation which starts at the underlying form  $\mathbf{b}$  and applies again rule  $\mathbb{A}$  first followed by rule  $\mathbb{B}$  yields a surface form  $\mathbf{c}$  which is different from  $\mathbf{b}$  under mild assumptions. In other words, the SPE phonology corresponding to the ordered rules  $\langle \mathbb{A}, \mathbb{B} \rangle$  yields the chain shift  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$ .  $\square$*

### 6.2.2. Counter-bleeding and saltations

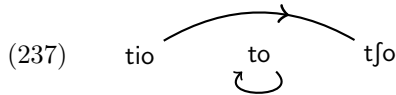
Consider a target derivation (232a) which takes an underlying form **a** and applies rule **A** first followed by rule **B**. Consider the derivation (232b) where the order of the two rules **A** and **B** is swapped. Rule **B** is said to *counter-bleed* rule **A** relative to the derivation (235a) provided rule **A** applies non-vacuously in that derivation but it applies vacuously in the swapped derivation (235b). In other words, **B** counter-bleeds **A** relative to the derivation (235a) if and only if **B** bleeds **A** relative to the swapped derivation (235b).

- (235) a. UR: **a**  
           rule **A**: NON VACUOUS  
           rule **B**: ...  
       b. UR: **a**  
           rule **B**: ...  
           rule **A**: VACUOUS

As an example (taken from Baković 2011), let **A** be a rule which palatalizes consonants before front vowels and let **B** be a rule which deletes the first vowel of a hiatus, exactly as in (233). Deletion counter-bleeds palatalization in the derivation (236a), as shown by comparison with the swapped derivation (236b).

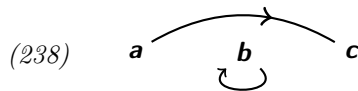
- (236) a. UR: /tio/  
           rule **A** = Palatalization: tʃio (NON VACUOUS)  
           rule **B** = Deletion: tʃo  
       b. UR: /tio/  
           rule **B** = Deletion: to  
           rule **A** = Palatalization: to (VACUOUS)

Counter-bleeding is said to yield *over-application opacity*: the palatalization rule **A** has over-applied in the derivation (236a), as the surface form [tʃo] obtained through that derivation displays the result of palatalization without a front vowel. This intuition can be restated as follows: the SPE phonology corresponding to the two ordered rules  $\langle \mathbf{A}, \mathbf{B} \rangle = \langle \text{Pal}, \text{Del} \rangle$  yields the saltation (237) whereby /tio/ shoots all the way to [tʃo] skipping over [to] which is “closer” in the sense that it involves less changes (only deletion rather than both deletion and palatalization).



As stated by the following proposition, this conclusion holds in general under mild assumptions: counter-bleeding opacity entails saltations. The “mild assumptions” mentioned in the informal statement of the proposition are made explicit in appendix A.2.

**Proposition 28.** *Let **b** the result of applying rule **B** to an underlying form **a**. Let **c** be the result of applying rule **A** first followed by rule **B** to that same underlying form **a**. If the latter derivation is a counter-bleeding derivation, then the SPE phonology corresponding to the ordered rules  $\langle \mathbf{A}, \mathbf{B} \rangle$  maps the underlying form **b** faithfully to itself. The resulting pattern*



*is a saltation because **a** is closer to **b** than to **c**, as **b** is obtained from **a** through rule **B** only while **c** is obtained from **a** through both rules **A** and **B** which are both non-vacuous under mild assumptions.* □

Note that the mappings depicted in (238) hold irrespectively of these additional assumptions. The latter assumptions are only needed in order for both rule **A** and **B** to be non-vacuous in the derivation from **a** to **c**, so that effectively **c** is further away from **a** than **b** is.

### 6.2.3. Output-drivenness, counter-feeding, and counter-bleeding

Phonological patterns are usually sorted into *opaque* and *transparent*. Opacity is defined in terms of rule ordering and equated with either counter-feeding or counter-bleeding rule orders. Assume that the

distinction between opacity and transparency is *psychologically real*, namely that it has cognitive implications for a number of phenomena related to historical sound change, acquisition, and artificial learning. The switch from rule-based to constraint-based phonology then requires a rule-independent definition of opacity. Tesar submits that output-drivenness provides such a rule-independent characterization of the transparent/opaque distinction:

$$(239) \quad \begin{array}{l} \text{transparent} \sim \text{output-driven} \\ \text{opaque} \sim \text{non-output-driven} \end{array}$$

The two propositions 27 and 28 easily lead to this connection between opacity and non-output-drivenness, as discussed in the rest of this subsection and summarized in table 1.

By proposition 27, counter-feeding opacity yields chain shifts (under mild assumptions). As seen in subsection 3.1.4, chain shifts are equivalent to the negation of idempotency. Non-idempotency thus captures half of the classical definition of opacity, namely counter-feeding opacity. The same holds for non-output-drivenness, as output-drivenness entails idempotency, as seen in section 5.1. It is worth pointing out that, nonetheless, the classical intuition of “what goes wrong” in counter-feeding opacity is turned upside down when we switch from rule interaction to idempotency. To illustrate, consider again the concrete example in (233)-(234). According to the classical rule-based intuition, what goes wrong in this case is under-application: the underlying form /tʰi/ *under-shoots* and ends up as [tʰi] instead of going all the way to [tʰj]. According to idempotency instead, what goes wrong is that /tʰi/ *over-shoots*: it goes all the way to [tʰj] when it should instead have stopped to the faithful [tʰi].

By proposition 28, counter-bleeding opacity yields saltations (under mild assumptions). As seen in subsection 5.1.3, saltations are not output-driven (modulo a careful definition of the similarity order). Non-Output-Drivenness thus captures counter-bleeding opacity. It is worth pointing out that, nonetheless, the classical intuition of what goes wrong in counter-bleeding opacity is turned upside down when we switch from rule interaction to output-drivenness. To illustrate, consider again the concrete example in (236)-(237). According to the classical rule-based intuition, what goes wrong in this case is over-application: the underlying form /tʰio/ *over-shoots* and ends up all the way at [tʰjo] instead of going just to [tʰo]. According to output-drivenness, what goes wrong is instead the fact that /tʰio/ *under-shoots*: it goes faithfully to [tʰo] itself when it should have instead mapped further away to [tʰjo].

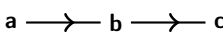
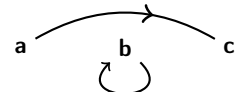
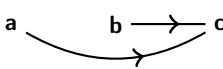
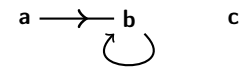
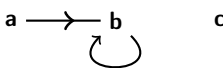
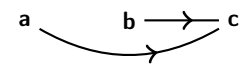
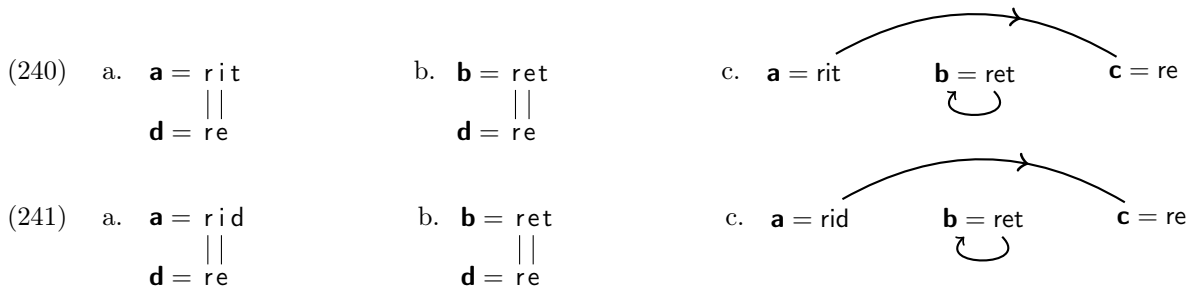
type of opacity:	counter-feeding	counter-bleeding
resulting phonological pattern:	a chain shift: 	a saltation: 
intuition of what goes wrong according to SPE:	<b>a</b> under-shoots ( <b>under-application opacity</b> ) and the following should have happened instead: 	<b>a</b> over-shoots ( <b>over-application opacity</b> ) and the following should have happened instead: 
intuition of what goes wrong according to output-drivenness:	<b>b</b> over-shoots and the following should have happened instead: since <b>b</b> attracts <b>a</b> then it should a fortiori attract <b>b</b> itself 	<b>b</b> under-shoots and the following should have happened instead: since <b>c</b> attracts <b>a</b> , then it should a fortiori attract <b>b</b> 

TABLE 1. Opacity according to SPE and output-drivenness

#### 6.2.4. Refinements

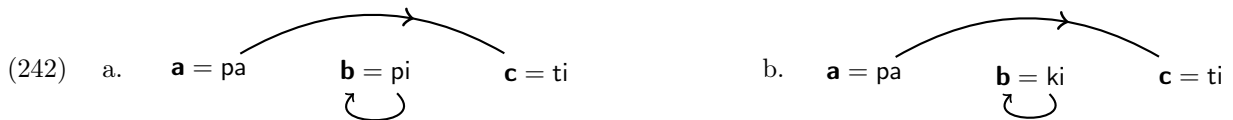
Whether a phonological pattern counts as a failure of output-drivenness effectively depends on the notion of phonological similarity used to define output-drivenness. In the preceding subsection, I have avoided this issue by illustrating the connection between counter-bleeding/saltations and non-output-drivenness in a particularly simple case where the relevant phonological forms consist of a single segment each and similarity is thus unambiguously defined in terms of feature value mismatches. Yet, we have seen in sections 5.3 and 5.4 that formalizing a notion of similarity adequate to output-drivenness is not a trivial task in the general case. In particular, we have compared two approaches: a weaker notion of similarity provided by definition 3 and axiomatized by definition 5; and a stronger notion of similarity provided by definition 4 and axiomatized by definition 6. The switch from the weak to the strong definition of similarity obviously has implications for what counts as a failure of output-drivenness, and thus for the connection between non-output-drivenness and opacity. The relationship between counter-bleeding opacity (or saltations) and output-drivenness thus requires a more careful discussion which takes into account the difference between these two approaches to similarity. This subsection only illustrates this point with a couple of examples

According to the weak notion of the similarity order provided by definitions 3 and 5, there is no difference between the two cases (240) and (241): in both cases, the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  counts as having less internal similarity than the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$ . Whether the two underlying deleted codas are identical as in (240a)/(240b) or different as in (241a)/(241b) is irrelevant. Hence, both patterns (240c) and (241c) count as a saltation and a failure of output-drivenness relative to the weak similarity order.



The situation is different relative to the strong notion of similarity order provided by definitions 4 and 6. Relative to the latter similarity order, the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  counts as having less internal similarity than the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  only in the case of (240), where the two underlying deleted codas are identical; but not in the case of (241), where the two candidates are simply not comparable relative to their internal similarity because of the two different underlying deleted codas. Hence, only the pattern (240c) but not the pattern (241c) counts as a saltation and a failure of output-drivenness relative to the strong similarity order.

Analogous considerations hold for (242): both patterns (242a) and (242b) count as a saltation and a failure of output-drivenness relative to the weak similarity order, while only the pattern (242a) but not the pattern (242b) qualifies relative to the strong similarity order.



Let me close by noting that the characterization of counter-bleeding opacity in terms of rule ordering fails at making any distinction between the two patterns (240c) and (241c) as well as between the two patterns (242a) and (242b). For instance, the pattern (241c) admits a straightforward counter-bleeding derivation. Consider the three rules in (243): rule  $\mathbb{A}$  deletes coronals after high vowels; rule  $\mathbb{B}$  lowers high vowels; and rule  $\mathbb{C}$  devoices final obstruents. Obviously, the underlying form /ret/ is faithfully mapped to [ret] by the SPE phonology  $\langle \mathbb{A}, \mathbb{B}, \mathbb{C} \rangle$ . Furthermore, the underlying form /rid/ is mapped to [re], as shown by the derivation (244a). The rule  $\mathbb{A}$  applies non-vacuously in the latter derivation. This is therefore a case of counter-bleeding, as the rule  $\mathbb{A}$  instead applies vacuously in the derivation (244b) where  $\mathbb{A}$  has been swapped with  $\langle \mathbb{B}, \mathbb{C} \rangle$ .

(243) a.  $\mathbb{A} : \text{t/d} \rightarrow \emptyset / i \_$

- b.  $\mathbb{B} : i \rightarrow e$   
c.  $\mathbb{C} : d \rightarrow t/_\#$
- (244) a.  $\text{rid} \xrightarrow{\mathbb{A}} \text{ri} \xrightarrow{\mathbb{B}} \text{re} \xrightarrow{\mathbb{C}} \text{re}$  ( $\mathbb{A}$  applies non-vacuously)  
b.  $\text{rid} \xrightarrow{\mathbb{B}} \text{red} \xrightarrow{\mathbb{C}} \text{ret} \xrightarrow{\mathbb{A}} \text{ret}$  ( $\mathbb{A}$  applies vacuously)

There is a potentially interesting idea here: that derivations with three rules (as opposed to derivations with only two rules traditionally considered in the rule-based theory of opacity) can be used to construct cases of counter-bleeding opacity which do not count as non-output-driven relative to the strong similarity order. A more detailed investigation of this conjecture is left for future research.

### 6.3. Output-drivenness or input-drivenness?

This section sketches an alternative to Tesar’s notion of output-drivenness, which I call *input-drivenness* because it adopts a *production* perspective (focused on the underlying or input forms) rather than Tesar’s *interpretation* perspective (focused on the surface or output forms). Like Tesar’s output-drivenness, also input-drivenness fails for chain shifts and saltations, so that its negation captures both counter-feeding opacity (chain shifts) and counter-bleeding opacity (saltations). Furthermore, like Tesar’s output-drivenness, also input-drivenness turns out to be related to the condition that the faithfulness constraints measure phonological distance in compliance with the metric triangle inequality. I conclude with some remarks on how to pull apart output-drivenness and input-drivenness.

#### 6.3.1. Input-drivenness

Tesar’s notion of output-drivenness introduced in section 5.1 focuses on two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  which share a surface form  $\mathbf{d}$  and differ for their underlying forms  $\mathbf{a}$  and  $\mathbf{b}$ . These two underlying forms are assumed to differ with respect to their similarity to the shared surface form:  $\mathbf{a}$  is less similar to  $\mathbf{d}$  than  $\mathbf{b}$  is, as formally captured by the similarity inequality (245). From now on, I will denote by  $\leq_{\text{sim}}^{\text{OD}}$  the similarity order used in the theory of Tesar’s output-drivenness.

$$(245) \quad (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\text{OD}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$$

Throughout this section, I explore an alternative notion of input-drivenness which focuses instead on two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}})$  and  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  which share the underlying form  $\mathbf{a}$  and differ for their surface forms  $\mathbf{c}$  and  $\mathbf{d}$ . These two surface forms are assumed to differ with respect to their similarity to the shared underlying form:  $\mathbf{d}$  is less similar to  $\mathbf{a}$  than  $\mathbf{c}$  is, as formally captured by the similarity inequality (246). I denote by  $\leq_{\text{sim}}^{\text{ID}}$  the similarity order used in the theory of this alternative notion of input-drivenness.

$$(246) \quad (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{b}, \mathbf{d}}) \leq_{\text{sim}}^{\text{ID}} (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$$

Tesar’s original notion of output-drivenness adopts the perspective of *interpretation*: given a surface form, how do I compute a corresponding underlying form? I am now interested in developing an analogous notion from the alternative perspective of *production*: given an underlying form, how do I compute a corresponding surface form?

Given the similarity inequality (246), suppose that a phonological grammar  $G$  declares the surface form  $\mathbf{c}$  phonotactically licit. This means that there exists some underlying form  $\mathbf{b}$  which  $G$  maps to  $\mathbf{c}$ , namely  $G(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ . Since the underlying form  $\mathbf{a}$  is more similar to  $\mathbf{c}$  than to  $\mathbf{d}$  and since  $\mathbf{c}$  is already phonotactically licit, the grammar  $G$  should not be able to map  $\mathbf{a}$  to  $\mathbf{d}$ , namely  $G(\mathbf{a}) \neq (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$ . Let me call any grammar  $G$  which abides by this logic *input-driven* relative to the similarity order  $\leq_{\text{sim}}^{\text{ID}}$ , yielding the following definition. The connection between the two implications (162) and (247) which define Tesar’s original output-drivenness and the alternative input-drivenness considered here is not immediately evident.

**Definition 7.** [Input-drivenness] A grammar  $G$  is *input-driven* relative to a similarity order  $\leq_{\text{sim}}^{\text{ID}}$  provided the following implication holds

$$(247) \quad \begin{array}{l} \text{If:} \quad G(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) \\ \text{Then:} \quad G(\mathbf{a}) \neq (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \end{array}$$

for any two different candidates  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  and  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\text{ID}} (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  and for any third candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ .

In other words, input-drivenness says that a grammar  $G$  maps any underlying form  $\mathbf{a}$  to a candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{d}})$  which has the highest internal similarity among the candidates which share that underlying form  $\mathbf{a}$  and have a surface form  $\mathbf{c}$  which is phonotactically licit relative to  $G$ , as stated in (248).

$$(248) \quad G(\mathbf{a}) \in \max_{\leq_{\text{sim}}^{\text{ID}}} \left\{ (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}}) \mid \mathbf{c} \text{ is phonotactically licit relative to } G \right\}$$

To illustrate, consider the special case where each underlying form admits a unique surface form which counts as the most similar one among the surface forms which are phonotactically licit. In this case, the phonotactics (together with the similarity order) completely determines the grammar  $G$ : input-drivenness requires  $G$  to map any underlying form to the most similar licit surface form.

### 6.3.2. Input-drivenness entails idempotency

As noted in subsection 5.1.2, it makes sense to assume that identity candidates have the greatest internal similarity. The similarity order  $\leq_{\text{sim}}^{\text{ID}}$  thus needs to satisfy condition (249) for any candidate  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c}, \mathbf{d}})$  and the corresponding identity candidate  $(\mathbf{c}, \mathbf{c}, \mathbb{I}_{\mathbf{c}, \mathbf{c}})$ .

$$(249) \quad (\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c}, \mathbf{d}}) \leq_{\text{sim}}^{\text{ID}} (\mathbf{c}, \mathbf{c}, \mathbb{I}_{\mathbf{c}, \mathbf{c}})$$

Whenever condition (249) holds, input-drivenness entails idempotency. In fact, input-drivenness requires the implication (247) to hold for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  such that the former has less internal similarity than the latter. Condition (249) ensures that is indeed the case when the two strings  $\mathbf{a}$  and  $\mathbf{c}$  coincide (and furthermore  $\rho_{\mathbf{a}, \mathbf{c}}$  is the identity correspondence relation on the string  $\mathbf{a} = \mathbf{c}$ ). In this case, the implication (247) in the definition of input-drivenness specializes to (250).

$$(250) \quad \begin{array}{l} \text{If:} \quad G(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) \\ \text{Then:} \quad G(\mathbf{c}) \neq (\mathbf{c}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \end{array}$$

Since the consequent of (250) holds for every candidate  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  different from  $(\mathbf{c}, \mathbf{c}, \mathbb{I}_{\mathbf{c}, \mathbf{c}})$ , I obtain (251), which is in turn the implication (52) in the definition of idempotency investigated in chapter 3.

$$(251) \quad \begin{array}{l} \text{If:} \quad G(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) \\ \text{Then:} \quad G(\mathbf{c}) = (\mathbf{c}, \mathbf{c}, \mathbb{I}_{\mathbf{c}, \mathbf{c}}) \end{array}$$

In conclusion, the definition of idempotency coincides with the definition of input-drivenness in the special case where  $\mathbf{a} = \mathbf{c}$  and  $\rho_{\mathbf{a}, \mathbf{c}}$  is the identity correspondence relation.

### 6.3.3. Input-drivenness and saltations

Consider the candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  and  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$  in (252). Any reasonable definition of the similarity order  $\leq_{\text{sim}}^{\text{ID}}$  should guarantee that the candidate  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}})$  has less internal similarity than the candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$ , since the former involves a disparity for two features (voicing and continuancy) while the latter involves a disparity for only one feature (continuancy, not voicing). Thus,  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a}, \mathbf{d}}) \leq_{\text{sim}}^{\text{ID}} (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{c}})$ .

$$(252) \quad \begin{array}{cc} \text{a. } \mathbf{a} = /p/ & \text{b. } \mathbf{a} = /p/ \\ \quad \quad \quad | & \quad \quad \quad | \\ \mathbf{d} = [\beta] & \mathbf{c} = [b] \end{array}$$

The phonological pattern (253) is therefore a *saltation*: the voiceless stop  $\mathbf{a}$  is mapped to the voiced fricative  $\mathbf{d}$ , skipping or jumping over the closer licit voiced stop  $\mathbf{c}$ .

$$(253) \quad \begin{array}{ccc} & \curvearrowright & \\ \mathbf{a} = p & & \mathbf{d} = \beta \\ & \curvearrowleft & \\ & \mathbf{c} = b & \end{array}$$

Input-drivenness crucially fails in the case of a saltation such as (253). In fact, the antecedent of the implication (247) which defines input-drivenness holds (with  $\mathbf{b} = \mathbf{c}$ ) but the consequent fails. Intuitively, the phonotactically licit surface form  $\mathbf{c}$  acts as a barrier which prevents  $\mathbf{a}$  to shoot all the way to  $\mathbf{d}$ .

### 6.3.4. Sufficient conditions for input-drivenness

Which conditions ensure that the OT grammars corresponding to any ranking of a given constraint set is input-driven? The answer to this question is provided by the following proposition 29. This proposition is completely analogous to Tesar’s proposition 18 from subsection 5.2.2, which provides sufficient constraint conditions for the original notion of output-drivenness. The proof of this proposition 29 is omitted because it is indeed completely analogous to Tesar’s proof of his proposition 18.

**Proposition 29.** *Assume that, for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  such that  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{sim}^{ID} (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ , for any other candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , the candidate set also contains a candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  such that the two following implications (254) hold for every faithfulness constraint  $F$  in the constraint set.*

$$(254) \quad \begin{array}{ll} a. & \text{If: } F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \\ & \text{Then: } F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \\ b. & \text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \\ & \text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) < F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \end{array}$$

Then, the OT grammar corresponding to any ranking of the constraint set is input-driven relative to the similarity order  $\leq_{sim}^{ID}$ , no matter what the markedness constraints look like.  $\square$

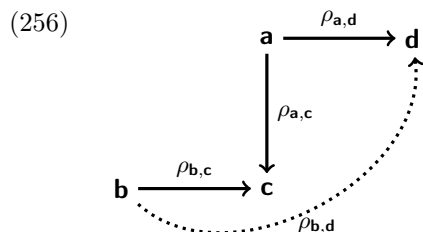
The assumption made by the proposition is twofold. First, it restricts the candidate set through the following condition: if two surface forms  $\mathbf{c}$  and  $\mathbf{d}$  share an underlying form  $\mathbf{a}$ , namely if the candidate set contains the two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ ; then, the surface forms  $\mathbf{c}$  and  $\mathbf{d}$  share every underlying form, namely the candidate set cannot contain a candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  without also containing the candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ . Second, the assumption of the proposition restricts the constraint set: it requires all the faithfulness constraints to satisfy the two implications (254). The latter two implications are identical to the two implications (170) which appear in Tesar’s original output-drivenness proposition 18 and have been referred to as the  $FODC_{comp}^{OT}$ . Despite involving the *same* two implications, the two propositions impose *different* faithfulness constraint conditions because they differ for the two candidates which have to stand in the similarity relation. I highlight this crucial difference in (255), where I have boxed the candidates which stand in a similarity relation in the two cases.

$$(255) \quad \begin{array}{ll} \text{Output-drivenness:} & \text{Input-drivenness:} \\ a. \text{ If: } \boxed{F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})} < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) & \text{If: } \boxed{F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})} < \boxed{F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})} \\ \text{Then: } \boxed{F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})} < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) & \text{Then: } F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < \boxed{F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})} \\ b. \text{ If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \boxed{F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})} & \text{If: } F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) \\ \text{Then: } F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) < \boxed{F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})} & \text{Then: } \boxed{F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})} < \boxed{F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})} \end{array}$$

The next subsection turns to the following question: which faithfulness constraints satisfy the sufficient condition for input-drivenness provided by proposition 29?

### 6.3.5. The triangle inequality entails input-drivenness

Let’s start with the second implication (254b) which appears in proposition 29. Consider three candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ ,  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ , and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  which are universally quantified over by proposition 29. Their correspondence relations are represented through solid arrows in (256). Proposition 29 requires in particular the existence of a fourth candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$  which puts the two strings  $\mathbf{b}$  and  $\mathbf{d}$  in correspondence through some relation  $\rho_{\mathbf{b},\mathbf{d}}$ , as represented by the dotted arrow in (256).





It is natural to assume that  $\rho_{\mathbf{b},\mathbf{d}}$  is the composition of: the correspondence relation  $\rho_{\mathbf{b},\mathbf{c}}$ , the inverse of the correspondence relation  $\rho_{\mathbf{a},\mathbf{c}}$  (which is denoted by  $\rho_{\mathbf{c},\mathbf{a}}$ ), and the correspondence relation  $\rho_{\mathbf{a},\mathbf{d}}$ . The existence of this composition correspondence relation is guaranteed by the transitivity axiom (66) from subsection 3.2.4. The implication (254b) can thus be specialized as in (257).

$$(257) \quad \begin{array}{l} \text{If:} \quad F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \overbrace{\rho_{\mathbf{b},\mathbf{c}}\rho_{\mathbf{c},\mathbf{a}}\rho_{\mathbf{a},\mathbf{d}}}^{\rho_{\mathbf{b},\mathbf{d}}}) \\ \text{Then:} \quad F(\mathbf{a}, \mathbf{c}) < F(\mathbf{a}, \mathbf{d}) \end{array}$$

Assume that the faithfulness constraint  $F$  satisfies the triangle inequality as formalized by the  $\text{FTI}_{\text{compcomp}}$  in section 4.1. By applying the  $\text{FTI}_{\text{comp}}$  to the two candidates  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$  and  $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{a}}\rho_{\mathbf{a},\mathbf{d}})$  and their composition candidate  $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{c}}\rho_{\mathbf{c},\mathbf{a}}\rho_{\mathbf{a},\mathbf{d}})$ , I obtain the inequality (258). The latter inequality in turn says that the antecedent of the (257) never holds, whereby the implication is trivially satisfied.

$$(258) \quad F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{c}}\rho_{\mathbf{c},\mathbf{a}}\rho_{\mathbf{a},\mathbf{d}}) \leq F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) + F(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{a}}\rho_{\mathbf{a},\mathbf{d}})$$

Let's now turn to the first implication (254a) which appears in proposition 29. In order to make further progress in the theory of input-drivenness, I need to be more explicit about the similarity order  $\leq_{\text{sim}}^{\text{ID}}$  that input-drivenness is predicated on. As noted in subsection 5.3.1, within constraint-based phonology, it is natural to assess similarity through the faithfulness constraints. Intuitively, a candidate has more internal similarity than another candidate provided the former incurs less faithfulness violations than the latter. This intuition can be formalized into the following axiom on the similarity order  $\leq_{\text{sim}}^{\text{ID}}$ : for any two candidates  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$  and  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  which share the underlying form  $\mathbf{a}$ , let  $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\text{ID}} (\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$  provided the inequality (259) holds for every faithfulness constraint  $F$  in the constraint set.<sup>1</sup>

$$(259) \quad F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$$

This axiom was found in subsection 5.3.2 to be too weak to ensure output-drivenness by itself, even for the simplest faithfulness constraint set, consisting of just the plain faithfulness constraints IDENT and MAX. This initial condition on similarity had therefore to be strengthened along the lines of subsection 5.3.3 in order to guarantee output-drivenness. The case of input-drivenness turns out to be different: the intuitive axiom (259) on the similarity order  $\leq_{\text{sim}}^{\text{ID}}$  turns out to suffice to ensure input-drivenness. In fact, this axiom (259) suffices to ensure that the antecedent in the first implication (254a) can never hold, whereby the implication is trivially satisfied.

### 6.3.6. How to pull apart output-drivenness and input-drivenness

Phonological patterns are usually sorted into *opaque* and *transparent*. Opacity is defined in terms of rule ordering and equated with either counter-feeding or counter-bleeding rule orders. Counter-feeding opacity yields chain shifts (subsection 6.2.1). Chain shifts are equivalent to the negation of idempotency (subsection 3.1.4). And both output-drivenness and input-drivenness entail idempotency (subsections 5.1.2 and 6.3.2). Hence, the negation of either output-drivenness or input-drivenness covers cases of counter-feeding opacity. Furthermore, counter-bleeding opacity yields saltations (subsection 6.2.2). And both output-drivenness and input-drivenness fail for saltations (subsections 5.1.3 and 6.3.3). Hence, the negation of either output-drivenness or input-drivenness covers cases of counter-bleeding opacity. In conclusion, we now have at our disposal the two notions of output-drivenness and input-drivenness which both capture the distinction between opacity and transparency at least in basic cases. How can output-drivenness and input-drivenness be pulled apart? This subsection offers some preliminary considerations.<sup>2</sup>

Let me take a closer look at a concrete case of saltation such as (260). Both output-drivenness and input-drivenness fail in this case. But the failure is derived in different ways in the two cases. Let me make the difference explicit.

$$(260) \quad \begin{array}{c} \text{p} \quad \text{v} \\ \quad \curvearrowright \\ \quad \text{f} \\ \quad \curvearrowleft \end{array}$$

<sup>1</sup> Obviously,  $\leq_{\text{sim}}^{\text{ID}}$  is reflexive and transitive. In order for it to also be antisymmetric (and thus qualify as a partial order), the set of faithfulness constraints must be rich enough that for any two candidates which share the underlying form, there exists at least one faithfulness constraint which assigns them a different number of violations.

<sup>2</sup> This subsection is based on discussion with Ezer Rasin.

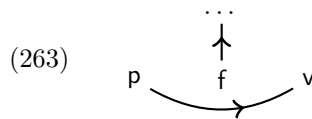
Tesar's output-drivenness compares the two candidates (261) which share the surface form [v] and differ for the two underlying forms /p/ and /f/. The reason why output-drivenness fails in the case of the saltation (260) can be informally stated as follows: since the surface form [v] attracts the underlying form /p/ from far away, then [v] should also attract any other underlying form closer to it, including in particular /f/.

$$(261) \quad (/p/, [v]) \leq_{\text{sim}}^{\text{OD}} (/f/, [v])$$

The alternative notion of input-drivenness explored in this section instead compares the two candidates (262) which share the underlying form /p/ and differ for the two surface forms [v] and [f]. The reason why input-drivenness fails in the case of the saltation (260) can be informally stated as follows: since the surface form [f] is phonotactically licit, it acts as a barrier and thus prevents the underlying form /p/ to shift any further into the surface form [v].

$$(262) \quad (/p/, [v]) \leq_{\text{sim}}^{\text{OD}} (/p/, [f])$$

Crucially, the two notions of output-drivenness make different predictions for a variant of the saltation pattern (260) illustrated in (263): the intermediate form /f/ is not faithfully realized but neutralized to something different than [v].



type of opacity:	counter-feeding	counter-bleeding
resulting phonological pattern:	a chain shift: $a \rightarrow b \rightarrow c$	a saltation: $a \rightarrow b \rightarrow c$ 
intuition of what goes wrong according to SPE:	<b>a</b> under-shoots ( <b>under-application opacity</b> ) and the following should have happened instead: $a \rightarrow b \rightarrow c$ 	<b>a</b> over-shoots ( <b>over-application opacity</b> ) and the following should have happened instead: $a \rightarrow b \rightarrow c$ 
intuition of what goes wrong according to output-drivenness:	<b>b</b> over-shoots and the following should have happened instead: since <b>b</b> attracts <b>a</b> then it should a fortiori attract <b>b</b> itself $a \rightarrow b \rightarrow c$ 	<b>b</b> under-shoots and the following should have happened instead: since <b>c</b> attracts <b>a</b> , then it should a fortiori attract <b>b</b> $a \rightarrow b \rightarrow c$ 
intuition of what goes wrong according to input-drivenness:	<b>b</b> over-shoots and the following should have happened instead: since <b>b</b> is phonotactically licit (as it is the surface realization of <b>a</b> ) it acts like a barrier and prevents <b>b</b> itself from shooting any further $a \rightarrow b \rightarrow c$ 	<b>a</b> over-shoots and the following should have happened instead: since <b>b</b> is phonotactically licit, it acts like a barrier and prevents <b>a</b> from shooting beyond <b>b</b> all the way to <b>c</b> $a \rightarrow b \rightarrow c$ 

TABLE 2. Opacity according to SPE, output-drivenness, and input-drivenness

Tesar’s output-drivenness does not distinguish between the two patterns (260) and (263), as it fails in the latter case as well. In fact, the notion of output-drivenness is based on the intuition that alternations are governed by *attractors*. Since /p/ is mapped to [v], then [v] counts as an attractor for any underlying form which is not further away than /p/. Thus, [v] must in particular attract to itself the underlying form /f/. In other words, output-drivenness really requires /f/ to go to [v], any other option for /f/ is ruled out, just as the original option (260) of a faithful mapping of /f/ to [f]. The case of input-drivenness is different: it does distinguish between (260) and (263), because it fails in the former case (as seen above) but not in the latter case. In fact, the notion of input-drivenness is based on the intuition that alternations are governed by *barriers*. Since [f] is phonotactically licit in the original saltation (260), then it acts as a barrier which prevents /p/ from going further away to [v]. But since [f] is not licit in the variant (263), it does not count as a barrier and /p/ is effectively allowed to slide all the way to [v]. I summarize these considerations by updating table 1 into table 2.

I can now go back to my initial question: which of the two notions of output-drivenness and input-drivenness provides the best characterization of the distinction between transparency and opacity? An experimental paradigm developed in White (2014) might be used to address this question. The logics of his main experiment (experiment 1) is summarized in (264). Subjects in the target condition are trained on the alternation /p/→[v], which changes both voicing and continuancy. Subjects in the control condition are instead trained on the alternation /b/→[v] which changes only continuancy but not voicing. At test phase, subjects are tested on the forms they have not been trained on, namely the forms /b/ and /f/ in the target condition and the forms /p/ and /f/ in the control condition. For each such form, subjects are asked to choose between the faithful mapping of that form to itself and the neutralization of that form to [v]. The dependent variable is the percentage of times that subjects choose the neutralization mapping over the faithful mapping. The results are reported in the bottom row of (264).

(264)	Target condition	Control condition
<b>training:</b>	subjects are trained on the alternation /p/→[v] which changes two features	subjects are trained on the alternation /b/→[v] which changes a single feature
<b>test:</b>	<ul style="list-style-type: none"> <li>• /b/→[v] preferred 70% of the time to /b/→[b]</li> <li>• /f/→[v] preferred 45% of the time to /f/→[f]</li> </ul>	<ul style="list-style-type: none"> <li>• /p/→[v] preferred 20% of the time to /p/→[p]</li> <li>• /f/→[v] preferred 15% of the time to /f/→[f]</li> </ul>

The underlying form /f/ is the crucial one, because subjects in both the target and the control conditions are tested on that underlying form. The main finding is that subjects are more likely to choose the neutralization mapping /f/→[v] in the target condition (45% of the times) than in the control condition (15% of the times). White explains this result as follows. *Ceteris paribus*, subjects prefer the faithful mapping /f/→[f] to the neutralization mapping /f/→[v]. Yet, positing the faithful mapping /f/→[f] in the target condition leads to the saltation (260). Positing the faithful mapping /f/→[f] in the control condition instead does not yield a saltation. The larger frequency of neutralizations /f/→[v] in the target than in the control condition thus reflects a bias against saltations.

What are the formal properties of saltations which make them dispreferred and thus drive White’s result? Either output-drivenness or input-drivenness represents a natural answer to this question. Subjects in White’s experiment were asked to choose between the neutralization mapping /f/→[v] and the faithful mapping /f/→[f]. Hence, both notions of output-drivenness and input-drivenness predict the experimental result, because both notions penalize the saltation (260) which would follow from the choice of the faithful mapping /f/→[f] in the target condition. Yet, consider a variant of White’s experiment where the neutralization mapping /f/→[v] is pitted not against the faithful mapping /f/→[f] but against some other neutralization mapping /f/→[...]. In the target condition, the former neutralization mapping /f/→[v] yields the original saltation (260) while the latter neutralization mapping /f/→[...] yields the pattern (263). Suppose that Tesar’s output-drivenness is the relevant notion of output-drivenness targeted by the bias in White’s subjects. Crucially, output-drivenness fails for both patterns (260) and (263). Thus, subjects in the target condition have no reasons to choose between /f/→[v] and /f/→[...], because both yield a failure of output-drivenness. The percentages in the target and the control conditions are thus expected to be similar because subjects have no reason to choose between /f/→[v] and /f/→[...] in neither the target nor the control condition. Suppose instead that it is input-drivenness which is targeted by the bias in White’s subjects. Crucially, input-drivenness fails for the original saltation (260) but not

for the variant (263). Thus, subjects in the target condition are expected to prefer /f/→[. . .] (which does not lead to a failure of input-drivenness) to /f/→[v] (which does lead to a failure of input-drivenness). The percentages in the target and the control conditions are thus expected to be different, because the bias against input-drivenness has something to say in the target condition (but obviously nothing to say in the control condition). This issue is left open for future research.

## 6.4. Idempotency and chain shifts

A phonological grammar is idempotent provided it realizes faithfully any phonotactically licit forms. Equivalently, provided it displays no chain shifts, as explained in subsection 3.1.4. Chain shifts have been widely documented in adult phonology (Łubowicz 2011; Moreton 2004a; Moreton and Smolensky 2002), child phonology (Velten 1943; Smith 1973; Macken 1980; Dinnsen and Barlow 1998; Cho and Lee 2000, 2003; Dinnsen et al. 2001; Jesney 2007), second language acquisition (Lee 2000; Jesney 2007), and delayed phonological acquisition (Dinnsen and Barlow 1998; Dinnsen et al. 2011). OT approaches to chain shifts can be sorted into two groups. Some approaches trade some of the assumptions of (classical) OT, reviewed in subsection 3.2.1. Such approaches include: sympathy theory (McCarthy 1999); output-output correspondence (Burzio 1998; Benua 2000); targeted constraints (Wilson 2001); turbidity (Goldrick 2001); anti-faithfulness constraints (Alderete 2001, 2008); comparative markedness (McCarthy 2003a); candidate chains (McCarthy 2007); stratal OT (Bermúdez-Otero 2007); and contrast preservation constraints (Łubowicz 2012). Other OT approaches to chain shifts are instead framed squarely within the classical architecture. This section provides an overview of the latter approaches from the formal perspective of the results established in sections 3.2-3.5.

### 6.4.1. Only a sufficient condition?

Consider an arbitrary (reflexive and transitive) candidate set, an arbitrary constraint set, and an arbitrary constraint ranking. Section 3.2 has established the  $FIC_{\text{comp}}$  (67) repeated in (265) as a *sufficient* condition for the idempotency of the corresponding OT grammar.

$$(265) \quad \text{If:} \quad F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}}) = 0 \\ \text{Then:} \quad F(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}}}_{\rho_{\mathbf{a}, \mathbf{c}}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$$

This statement contains three universal quantifications: over candidate sets, over constraint sets, and over rankings. At this level of generality, the  $FIC_{\text{comp}}$  is not only a sufficient but also a *necessary* condition for idempotency, in the following sense. Consider a faithfulness constraint  $F$  which does not satisfy the  $FIC_{\text{comp}}$  (265). This means that there exist two candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$  such that:  $F$  assigns no violations to the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ , so that the antecedent of the  $FIC_{\text{comp}}$  holds; but  $F$  assigns more violations to the candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$  than to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , so that the consequent fails. Suppose that the constraint set also contains a markedness constraint  $M$  which assigns more violations to the surface form  $\mathbf{b}$  than to the surface form  $\mathbf{c}$ .<sup>3</sup> The OT grammar corresponding to the ranking  $F \gg M$  displays the chain shift  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  and thus fails at idempotency: the string  $\mathbf{b}$  is phonotactically licit, because the underlying form  $\mathbf{a}$  is mapped to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , as shown in (266a); yet, the string  $\mathbf{b}$  does not surface faithfully, because the underlying form  $\mathbf{b}$  is not mapped to the identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , as shown in (266b).

(266) a.  $F$  fails at the consequent of the  $FIC_{\text{comp}}$  (265):

$\mathbf{a}$	$F$	$M$
$\mathbb{I}_{\mathbb{E}}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$	* . . . *	*
$(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$	* . . . * !	

<sup>3</sup> Furthermore, assume that there exists a markedness constraint (either  $M$  or a different markedness constraint ranked above  $M$ ) which assigns more violations to the surface form  $\mathbf{a}$  than to the surface form  $\mathbf{b}$ . The latter markedness constraint is responsible for ruling out the candidate  $(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a}, \mathbf{a}})$ , which is therefore ignored in the rest of this section.

b.  $F$  satisfies the antecedent of the  $\text{FIC}_{\text{comp}}$  (265):

<b>b</b>	$F$	$M$
$(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$		*
$\mathbb{E}(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$		

The rest of this section shows how various approaches to chain shifts in the classical OT literature fit within the schema (266), where  $F$  is one of the faithfulness constraints which were shown in sections 3.3-3.5 to fail the  $\text{FIC}_{\text{comp}}$ .

#### 6.4.2. Chain-shifts through constraint conjunction

As noted in subsection 3.5.6, constraint conjunction yields faithfulness constraints which fail the  $\text{FIC}_{\text{comp}}$ . The use of constraint conjunction to model chain shifts within classical OT has been pioneered by Kirchner (1996) and systematized by Moreton and Smolensky (2002). To illustrate, consider the chain shift in the Lena dialect of Spanish in (267) (data from Hualde 1989 and Gnanadesikan 1997, section 5.4.3): because of the high final vowel in the masculine form, the underlying low and mid vowels (as revealed by the feminine form) raise to mid and high vowels, respectively.

(267)	$/a/ \rightarrow [e]$	<b>gáta</b> ‘cat-FEM’	<b>gétu</b> ‘cat-MAS’
		<b>sánta</b> ‘saint-FEM’	<b>séntu</b> ‘saint-MAS’
	$/e/ \rightarrow [i]$	<b>néna</b> ‘child-FEM’	<b>nínu</b> ‘child-MAS’
		<b>séks</b> ‘dry-FEM’	<b>síku</b> ‘dry-MAS’

The markedness constraint RAISE favors higher vowels before a high vowel. The conjunction of the two faithfulness constraints  $\text{IDENT}_{[\text{high}]}$  and  $\text{IDENT}_{[\text{low}]}$  punishes underlying low vowels mapped to surface high vowels. The analysis (268) is an instance of the scheme (266), with the conjoined constraint playing the role of the non- $\text{FIC}_{\text{comp}}$  faithfulness constraint.

(268) a.

$/gátu/$	$\text{ID}_{[\text{high}]} \wedge \text{ID}_{[\text{low}]}$ RAISE	
$\mathbb{E}[gétu]$		*
$[gítu]$	*!	

b.

$/nénu/$	$\text{ID}_{[\text{high}]} \wedge \text{ID}_{[\text{low}]}$ RAISE	
$[nénu]$		*!
$\mathbb{E}[nínu]$		

Other approaches proposed in the literature are equivalent to the approach based on the conjunction of identity faithfulness constraints (see also the discussion of V-HEIGHTDISTANCE in Kirchner 1995). For instance, Gnanadesikan (1997, chapter 3) accounts for the chain shift  $p \rightarrow b \rightarrow m$  (in post-nasal position) through the faithfulness constraint  $\text{IDENT-Adj}$ , which is violated by a voiceless obstruent and a corresponding sonorant because they are separated by a distance larger than 2 on the *inherent voicing scale*. This constraint is thus equivalent to the conjunction  $\text{IDENT}_{[\text{voice}]} \wedge \text{IDENT}_{[\text{son}]}$ . Analogously, Dinnsen and Barlow (1998) account for the chain shift  $s \rightarrow \theta \rightarrow f$  through the faithfulness constraint  $\text{DISTFAITH}$ , which is violated when the underlying and surface form differ by more than 1 on the scale  $f = 1$ ,  $\theta = 2$ , and  $s = 3$  and is thus equivalent to the conjunction  $\text{IDENT}_{[\text{coronal}]} \wedge \text{IDENT}_{[\text{strident}]}$ .

Yet, the approach based on constraint conjunction is more general than the latter approaches based on “scales”, as the former but not the latter extends to chain shifts which involve deletion (Moreton and Smolensky 2002). To illustrate, consider the Sea Dayak chain shift in (269) (data from Kenstowicz and Kisseberth 1979).

(269)	$/ŋga/ \rightarrow [ŋa]:$	$/naŋga/$	$[nãŋa?]$	‘set up a ladder’
	$/ŋa/ \rightarrow [ŋã]:$	$/naŋa/$	$[nãŋã?]$	‘straighten’

The analysis based on constraint conjunction extends as in (268) (based on Łubowicz 2011), which is another instance of the scheme (266).

(270) a.

$/naŋga/$	$\text{ID}_{[\text{nas}]} \wedge \text{MAX}$	*NV
$\mathbb{E}[nãŋa?]$		*
$[nãŋã?]$	*!	

b.

$/naŋa/$	$\text{ID}_{[\text{nas}]} \wedge \text{MAX}$	*NV
$[nãŋa?]$		*!
$\mathbb{E}[nãŋã?]$		

### 6.4.3. Chain shifts through constraint restrictions

As noted in section 3.4, constraint restriction yields faithfulness constraints which fail the  $\text{FIC}_{\text{comp}}$  when the correspondence relations are allowed to “cross” the restriction. This observation systematizes various approaches to chain shifts proposed in the classical OT literature. This subsection discusses some examples in detail, with the goal of showing that restricted faithfulness constraints actually offer a straightforward translation of the classical counter-feeding rule ordering of chain shifts.

#### 6.4.3.1. First example

Jesney (2005, 2007) considers the classical child chain shift in (271): coronal stridents are stopped; but underlying coronal stops are velarized when followed by a lateral (data from Amahl at age 2;2-2;11 as described in Smith 1973).

(271)	$/s, z, \int, \mathfrak{z}, \mathfrak{t}\mathfrak{f}, \mathfrak{d}\mathfrak{z}/ \rightarrow [t, d]:$	$[\text{p}\Lambda\text{d}\text{əl}]$ ‘puzzle’	$[\text{p}\text{a}:\text{t}\text{l}\text{i}:]$ ‘parsley’
		$[\text{p}\text{e}\text{t}\text{əl}]$ ‘special’	$[\text{æ}\text{n}\text{d}\text{ə}\text{l}\text{ə}]$ ‘Angela’
	$/t, d, n/ \rightarrow [k, g, \eta]:$	$[\text{p}\Lambda\text{g}\text{əl}]$ ‘puddle’	$[\text{b}\Lambda\text{k}\text{l}\text{ə}]$ ‘butler’
		$[\text{t}\text{æ}\eta\text{g}\text{əl}]$ ‘sandal’	$[\text{b}\text{ɔ}\text{k}\text{əl}]$ ‘bottle’

Jesney’s analysis is summarized in (272). It relies on what she calls the “specific” faithfulness constraint  $\text{IDENT}_{\text{Coronal}}/[+\text{strident}]$ , which mandates that input stridents preserve their coronality. This constraint can be re-interpreted as  $\text{IDENT}_{\varphi,R}$  where  $\varphi$  is the feature [coronal] and the restriction  $R$  pairs a string with the set of its stridents. As noted in subsection 3.4.5,  $\text{IDENT}_{\varphi,R}$  fails at the  $\text{FIC}_{\text{comp}}$  when correspondence relations are allowed to exit from the restriction  $R$  without changing the value of the feature  $\varphi$ , namely to put in correspondence an underlying segment which satisfies the restriction with a surface segment which does not and yet has the same value for the feature  $\varphi$ , as described in (83). That is precisely the case in (272), as the underlying  $/s/$  (which satisfies the restriction  $R$ ) corresponds to surface  $[t]$  (which does not satisfy the restriction) and yet they are both coronals.

(272)	a.	<table style="border-collapse: collapse; text-align: center;"> <tr> <td style="padding: 2px 10px;"><math>/s/</math></td> <td style="padding: 2px 10px;"><math>\text{IDENT}_{[\text{cor}]}/[+\text{str}]</math></td> <td style="padding: 2px 10px;">*TL</td> </tr> <tr> <td style="padding: 2px 10px;">☞ <math>[t]</math></td> <td></td> <td style="padding: 2px 10px;">*</td> </tr> <tr> <td style="padding: 2px 10px;"><math>[k]</math></td> <td style="padding: 2px 10px;">*!</td> <td></td> </tr> </table>	$/s/$	$\text{IDENT}_{[\text{cor}]}/[+\text{str}]$	*TL	☞ $[t]$		*	$[k]$	*!		b.	<table style="border-collapse: collapse; text-align: center;"> <tr> <td style="padding: 2px 10px;"><math>/t/</math></td> <td style="padding: 2px 10px;"><math>\text{IDENT}_{[\text{cor}]}/[+\text{str}]</math></td> <td style="padding: 2px 10px;">*TL</td> </tr> <tr> <td style="padding: 2px 10px;"><math>[t]</math></td> <td></td> <td style="padding: 2px 10px;">*!</td> </tr> <tr> <td style="padding: 2px 10px;">☞ <math>[k]</math></td> <td></td> <td></td> </tr> </table>	$/t/$	$\text{IDENT}_{[\text{cor}]}/[+\text{str}]$	*TL	$[t]$		*!	☞ $[k]$		
$/s/$	$\text{IDENT}_{[\text{cor}]}/[+\text{str}]$	*TL																				
☞ $[t]$		*																				
$[k]$	*!																					
$/t/$	$\text{IDENT}_{[\text{cor}]}/[+\text{str}]$	*TL																				
$[t]$		*!																				
☞ $[k]$																						

Jesney’s analysis (272) is thus an instance of the scheme (266), with the restricted constraint  $\text{IDENT}_{[\text{cor}],R}$  playing the role of the non- $\text{FIC}_{\text{comp}}$  faithfulness constraint.<sup>4</sup>

It is interesting to compare Jesney’s OT analysis with a rule-based analysis. Obviously, two rules are needed in order to derive the chain shift (271), corresponding to the two processes of velarization and stopping. Crucially, velarization must be ordered before stopping: there would otherwise be no difference between an underlying  $/t/$  and a  $t$  derived through stopping at the point when velarization applies, as shown by the two derivations in (273).

(273)	a.	underlying representation: $/s/$	b.	underlying representation: $/t/$
		stopping: $t$		stopping: $t$
		velarization: $[k]$		velarization: $[k]$

Yet, properly ordering the two rules does not suffice to get the chain shift (271). In fact, if velarization targeted all coronals (independently of their stridency or continuancy), velarization would incorrectly affect  $/s/$  as well, as shown by the derivations in (274). Thus, velarization must apply first and furthermore be restricted to non-tridents (or non-fricatives).

(274)	a.	underlying representation: $/s/$	b.	underlying representation: $/t/$
		velarization: $\times$		velarization: $k$
		stopping: $[k]$		stopping: $[k]$

In conclusion, the SPE analysis of the chain shift (271) can be made explicit as in (275). Here, I use a non-standard notation for SPE rules, where I pool together the conditions on the quality of the target (usually written on the left of the SPE rule formula) and the conditions on the environment (usually written on the right). The reason for that is that I want to draw a parallel with restricted faithfulness constraints, where the restriction can be defined either in terms of segmental quality or position.

<sup>4</sup> Assume that only coronals can be [+strident], while all non-coronals are [-strident]. Jesney’s constraint  $\text{IDENT}_{\text{Coronal}}/[+\text{strident}]$  is then provably equivalent to the conjoined constraint  $\text{IDENT}_{[\text{strident}]} \wedge \text{IDENT}_{[\text{coronal}]}$ .

- (275) a. First rule to apply: *restricted velarization rule*:  
 change: [place=cor]  $\rightarrow$  [place=dor]  
 quality conditions: the target is [-strid]  
 environment conditions: the target is followed by a lateral
- b. Second rule to apply: *stopping rule*:  
 change: [-cont]  $\rightarrow$  [+cont]  
 quality conditions: the target is [-son]  
 environment conditions: —

Interestingly, Jesney’s OT analysis (272) can be interpreted as a straightforward and faithful translation of the SPE analysis (275) into OT, as made explicit in (276). Each rule of the SPE analysis corresponds to a pair of a markedness and a faithfulness constraint. The restriction of the velarization rule (275a) to non-stridents corresponds to the restriction of the corresponding faithfulness constraint IDENT<sub>[place]</sub>/ [+str] to stridents. In other words, the faithfulness constraint is restricted to the *complement* of the quality condition of the corresponding SPE rule. The switch to the complement is intuitively due to the fact that SPE rules “do things” while faithfulness constraints “prevent things” from happening. The fact that it is the quality condition which matters for the restriction of the faithfulness constraint (rather than the environment restriction) captures the fact that the SPE analysis (275) is an instance of counter-feeding *on the focus*. Finally, the ordering of velarization before stopping in the SPE analysis (275) corresponds to the fact that the markedness constraint \*FRICATIVE corresponding to stopping (the second rule) must be ranked above the markedness constraint \*TL corresponding to velarization (the first rule), in order for /s/ to go to [t] rather than to stay at /s/.

(276)	SPE analysis	OT analysis
	<i>restricted velarization rule</i> (275a)	$\Rightarrow$ <i>restricted faithfulness</i> : IDENT <sub>[place]</sub> / [+str] markedness: *TL
	<i>stopping rule</i> (275b)	$\Rightarrow$ <i>faithfulness</i> : IDENT <sub>[cont]</sub> markedness: *FRICATIVE
	<i>velarization rule precedes stopping rule</i>	$\Rightarrow$ *FRICATIVE $\gg$ *TL

#### 6.4.3.2. Second example

Orgun (1995) considers the chain shift in (277) from Bedouin Hijazi Arabic: /a/ is raised to [i] but /i/ is deleted (both processes are restricted to short vowels in non-final open syllables; the data come from McCarthy 1993).

(277)	/a/ $\rightarrow$ [i]:	/katab/	[kitab]	‘he wrote’
		/rafaagah/	[rifaagah]	‘companions’
	/i/ $\rightarrow$ $\emptyset$ :	/ʕarif+at/	[ʕarfat]	‘she knew’
		/kitil/	[ktil]	‘he was killed’

Orgun’s analysis is summarized in (278)—plus a markedness constraint [\*a] which is omitted here. It relies on his constraint CORRESPOND(/a/), which mandates that “every input /a/ has an output correspondent”. This constraint can be re-interpreted as MAX<sub>R</sub> where the restriction *R* pairs a string with the set of its a’s. As shown in subsection 3.4.1, MAX<sub>R</sub> fails at the FIC<sub>comp</sub> when correspondence relations are allowed to exit from the restriction *R*, namely to put in correspondence an underlying segment which satisfies the restriction with a surface segment which does not, as described in (79). That is precisely the case in (278), as the underlying /a/ (which satisfies the restriction *R*) corresponds to the surface [i] (which does not satisfy the restriction). Orgun’s analysis (278) is thus an instance of the scheme (266), with the restricted constraint MAX<sub>R</sub> playing the role of the non-FIC<sub>comp</sub> faithfulness constraint.

(278) a.	<table border="1"> <thead> <tr> <th>/a/</th> <th>CORR(/a/)</th> <th>*V</th> </tr> </thead> <tbody> <tr> <td>☞ [i]</td> <td></td> <td>*</td> </tr> <tr> <td>☞ <math>\emptyset</math></td> <td>*!</td> <td></td> </tr> </tbody> </table>	/a/	CORR(/a/)	*V	☞ [i]		*	☞ $\emptyset$	*!		b.	<table border="1"> <thead> <tr> <th>/i/</th> <th>CORR(/a/)</th> <th>*V</th> </tr> </thead> <tbody> <tr> <td>☞ [i]</td> <td></td> <td>*!</td> </tr> <tr> <td>☞ <math>\emptyset</math></td> <td></td> <td></td> </tr> </tbody> </table>	/i/	CORR(/a/)	*V	☞ [i]		*!	☞ $\emptyset$		
/a/	CORR(/a/)	*V																			
☞ [i]		*																			
☞ $\emptyset$	*!																				
/i/	CORR(/a/)	*V																			
☞ [i]		*!																			
☞ $\emptyset$																					

It is interesting to compare Orgun’s OT analysis with a rule-based analysis. That analysis requires two rules corresponding to deletion and raising. Two crucial conditions must be met in order for the SPE analysis to succeed. The first condition is that deletion be ordered before raising. That by itself does

not suffice. An additional necessary condition is that the deletion rule be restricted to high (or non-low) vowels. The SPE analysis can thus be made explicit as in (279).


- (279) a. First rule to apply: *restricted deletion rule*:  
           change: deletion  
           quality conditions: the target is [+high]  
           environment conditions: the target is not the last vowel  
   the target belongs to an open syllable
- b. Second rule to apply: *raising rule*:  
           change: [+low, -high]  $\rightarrow$  [-low, +high]  
           quality conditions: —  
           environment conditions: —

Orgun's OT analysis (278) faithfully translates the SPE analysis (279), as made explicit in (280). Each rule of the SPE analysis corresponds to a pair of a markedness and a faithfulness constraint. The restriction of the deletion rule (279a) to high vowels corresponds to the restriction of the corresponding faithfulness constraint MAX/[+low, -high] to low vowels. In other words, the faithfulness constraint is restricted to the *complement* of the quality condition of the corresponding SPE rule. The fact that it is the quality condition which matters for the restriction of the faithfulness constraint (rather than the environment restriction) captures the fact that the SPE analysis (279) is an instance of counter-feeding *on the focus*.

(280)	SPE analysis	OT analysis
	<i>restricted</i> velarization rule $\Rightarrow$	<i>restricted</i> faithfulness: MAX/[+low, -high] markedness: *HIGHNONFINALVOWELOPENSYLLABLE
	stopping rule $\Rightarrow$	faithfulness: IDENT <sub>[low]</sub> , IDENT <sub>[high]</sub> markedness: RAISE

#### 6.4.3.3. Third example

McCarthy (2007, p. 34-36) considers the following chain shift from Bedouin Arabic: the vowel /a/ of /gabur/ belongs to a non-final open syllable and is therefore raised, yielding [gibur]; the vowel /a/ of /gabr/ does not belong to a non-final open syllable and is therefore not raised despite the fact that vowel epenthesis opens that syllable, yielding [gabur].

(281)    gabr  $\rightarrow$  gabur  $\rightarrow$  gibur 

The rule-based analysis of this chain shift requires two rules corresponding to raising and epenthesis. Two crucial conditions must be met in order for the SPE analysis to succeed. The first condition is that raising be ordered before epenthesis. That by itself does not suffice. An additional necessary condition is that raising be restricted to vowels in non-final open syllables. The SPE analysis can thus be made explicit as in (282).

- (282) a. First rule to apply: *restricted raising rule*:  
           change: [+low, -high]  $\rightarrow$  [-low, +high]  
           quality conditions: —  
           environment conditions: the target belongs to an open non-final syllable
- b. Second rule to apply: *epenthesis rule*:  
           change:  $\emptyset \rightarrow$  V  
           quality conditions: —  
           environment conditions: C\_C

McCarthy entertains an OT analysis based on the conjunction of DEP and IDENT<sub>[low]</sub> (within the domain of a single underlying syllable or two adjacent surface syllables). And he criticizes this analysis based on the fact that it fails to capture the following asymmetry: although raising does not happen when



epenthesis targets the complex coda *following* the vowel, it does happen when epenthesis targets the complex onset *preceding* the vowel, as shown by alternations such as /t<sup>ʃ</sup>arad ʌnam-ih/ → [t<sup>ʃ</sup>a.ra.diʌ.əni.mih]. The complete pattern thus looks as follows:

- (283) a. /Ca.σ/ → [Ci.σ]  
           /CaC.(σ)/ → [CaC.(σ)]  
       b. /CaCC.(σ)/ → [Ca.CVC.(σ)]  
           /CCa.σ/ → [CV.Ci.σ]

McCarthy's critique of the OT analysis seems to me disingenuous, as the context-free faithfulness constraint IDENT<sub>[low]</sub> he entertains actually fails to capture even the transparent facts in (283a), which have nothing to do with opacity. In order to get the pattern (283a), we need a positional variant of IDENT<sub>[low]</sub> which is restricted to the complement of the environment condition of the raising rule. That complement boils down to the condition that either the vowel is hosted by the final syllable or it is hosted by a closed syllable. Equivalently, either it is hosted by a final syllable or it is hosted by a non-final closed syllable. A vowel hosted by a non-final closed syllable can also be characterized as a vowel followed by a cluster C<sub>1</sub>C<sub>2</sub>. The SPE analysis (282) could therefore be restated in OT as in (284).

(284)	SPE analysis	OT analysis
	<i>restricted</i> raising rule ⇒	<i>restricted</i> faithfulness: IDENT <sub>[low]</sub> /last-vowel, IDENT <sub>[low]</sub> /_CC markedness: *[+low]
	epenthesis rule ⇒	faithfulness: DEP markedness: *CC

#### 6.4.4. Chain-shifts through breaking

Let me close this section by discussing a fictional example. Kubozono et al. (2008) report that English *frog* is imported as [fu.róg.gu] into Japanese: the velar stop geminates (despite being voiced) because of a requirement on the placement of stress, captured here through a place-holder constraint STRESS. Assume an analysis of consonant gemination in terms of breaking of a single underlying consonant into two surface copies, as indicated by the correspondence relations in (285). Subsection 3.3.3 has shown that plain identity faithfulness constraints fail at the FIC<sub>comp</sub> when the correspondence relations are allowed to break underlying segments. This fact could be used to derive a fictional chain shift such as η → g → gg through the analysis (285), which is an instance of the scheme (266) with the identity constraint playing the role of the non-FIC<sub>comp</sub> faithfulness constraint.

(285) a.	/ŋ/	IDENT <sub>[nas]</sub>	STRESS	b.	/g/	IDENT <sub>[nas]</sub>	STRESS
	↗ (ŋ, g) └──┘	*	*		(g, g)		*!
	(ŋ, gg) └──┘	**!			↗ (g, gg) └──┘		

This example is fictional because I have not been able to find a realistic case of chain shift which involves an underlying segment broken into two surface segments.

#### 6.4.5. Summary

Section 3.2 has shown that chain shifts require the FIC<sub>comp</sub> to fail. Based on sections 3.3-3.5, there are three major ways for the FIC<sub>comp</sub> to fail. One option is to use a faithfulness constraint which flouts the FIC<sub>comp</sub>, such as a faithfulness constraint obtained through constraint conjunction. A second option is to use the restriction of a faithfulness constraint which would otherwise comply with the FIC<sub>comp</sub>. The third option is to let the correspondence relations break underlying segments. The former two options have been exploited in the literature on chain shifts, as reviewed in this section.

### 6.5. Idempotency and the early acquisition of phonotactics

This section investigates the learnability implications of the structure provided by idempotency from the perspective of modeling the early stage of the acquisition of phonotactics.

### 6.5.1. Learning without underlying forms

The initial, most basic formulation of the language learning problem within constraint-based phonology is (286). Sticking with the generative perspective, the learner is provided with full typological information, by (286a.i). Since a grammar is effectively a set of candidates, the linguistic data that the learner is trained on are indeed a certain number  $\ell$  of candidates, by (286a.ii). The target grammar which has allegedly generated the data ought to be consistent with it. The learner is thus required to find a grammar in the typology which satisfies the consistency condition (286b).

- (286) a. *Given:* i. a typology specified through a candidate and a constraint set;  
 ii. a dataset consisting of  $\ell$  candidates  $(\mathbf{a}_1, \mathbf{b}_1, \rho_{\mathbf{a}_1, \mathbf{b}_1}), \dots, (\mathbf{a}_\ell, \mathbf{b}_\ell, \rho_{\mathbf{a}_\ell, \mathbf{b}_\ell})$ .  
 b. *Find:* a grammar  $G$  in the typology which is *consistent* with the data,  
 i.e.,  $G(\mathbf{a}_i) = (\mathbf{a}_i, \mathbf{b}_i, \rho_{\mathbf{a}_i, \mathbf{b}_i})$  for every  $i = 1, \dots, \ell$ .

We know how to solve efficiently this problem within both OT and HG (Tesar and Smolensky 1998; Potts et al. 2010), at least if the learner is allowed the resources to list all the candidates which share a given underlying form. Yet, this formulation (286) of the grammar learning problem suffers from two drawbacks: the success condition (286b) is too weak and the training data (286a.ii) are too rich, as discussed in the rest of this subsection.

The problem (286) is *ill-posed* in particular because it admits multiple solutions: how should the learner choose among the plausibly many grammars which all satisfy the consistency condition (286b)? Each grammar  $G$  generates the corresponding set  $L_G$  of surface forms which are phonotactically licit according to  $G$ . A natural additional condition on the grammar returned by the learner is that the corresponding set of licit surface forms be as small as possible compatibly with consistency, as stated by the added condition (287b.ii). This is the classical *Subset Principle* (Fodor and Sakas 2005).

- (287) a. *Given:* i. a typology specified through a candidate and a constraint set;  
 ii. a dataset consisting of  $\ell$  candidates  $(\mathbf{a}_1, \mathbf{b}_1, \rho_{\mathbf{a}_1, \mathbf{b}_1}), \dots, (\mathbf{a}_\ell, \mathbf{b}_\ell, \rho_{\mathbf{a}_\ell, \mathbf{b}_\ell})$ .  
 b. *Find:* a grammar  $G$  in the typology which is  
 i. *consistent* with the data,  
 i.e.,  $G(\mathbf{a}_i) = (\mathbf{a}_i, \mathbf{b}_i, \rho_{\mathbf{a}_i, \mathbf{b}_i})$  for every  $i = 1, \dots, \ell$ ;  
 ii. *restrictive* relative to the data,  
 i.e., no grammar  $\hat{G}$  such that  $L_{\hat{G}} \subsetneq L_G$  satisfies the consistency condition (287b.i).

The added restrictiveness condition (287b.ii) changes drastically the nature of the learning problem: while the original learning problem (286) is efficiently solvable in OT, the amended problem (287) is intractable (Magri 2013), namely it admits no efficient solution algorithm. Intractability holds even if the learner is allowed the resources to list and inspect all the candidates which share a given underlying form. This intractability result plausibly extends from OT to comparable frameworks such as HG (indeed Clark and Lappin 2011, p. 137 submit that “complexity problems are largely independent of the learning paradigm, as all frameworks encounter them.”).

Another shortcoming shared by the problem formulations (286) and (287) is that they both assume the linguistic data available to the learner to consist not only of the surface forms  $\mathbf{b}_1, \dots, \mathbf{b}_\ell$  but of the corresponding underlying forms  $\mathbf{a}_1, \dots, \mathbf{a}_\ell$  and correspondence relations  $\rho_{\mathbf{a}_1, \mathbf{b}_1}, \dots, \rho_{\mathbf{a}_\ell, \mathbf{b}_\ell}$  as well. This assumption is problematic because underlying forms (and correspondence relations) are most likely *hidden* to the learner in a realistic learning setting and indeed need to be inferred as part of the learning problem. This assumption that underlying forms are available to the learner is particularly implausible for what Hayes (2004) calls the *early stage* of the acquisition of phonology. This acquisition stage (around nine months) is characterized by two properties. The first property is that the child reacts differently to licit and illicit sound structures. In other words, the phonological grammar learned by the end of this stage satisfies both the consistency condition (287b.i) (whereby the learned grammar recognizes licit forms as such) and the restrictiveness condition (287b.ii) (whereby the learned grammar recognizes illicit forms as such). The second property of the early acquisition stage considered is that morphology is plausibly still lagging behind and the child has therefore no access to paradigms of morphological alternations. Lack of morphological paradigms plausibly means that the child is blind to alternations and thus has no information on the underlying forms. These considerations suggest a further reformulation of the learning problem, where the linguistic data available to the learner consist more realistically of nothing more than a certain number  $\ell$  of surface forms, as in (288a.ii).

- (288) a. *Given*: i. a typology specified through a candidate and a constraint set;  
 ii. a dataset consisting of  $\ell$  surface forms  $\mathbf{b}_1, \dots, \mathbf{b}_\ell$ .
- b. *Find*: a grammar  $G$  in the typology which is  
 i. *consistent* with the data,  
 i.e.,  $G(\mathbf{a}_i) = (\mathbf{a}_i, \mathbf{b}_i, \rho_{\mathbf{a}_i, \mathbf{b}_i})$  for some  $\mathbf{a}_i, \rho_{\mathbf{a}_i, \mathbf{b}_i}$ , for every  $i = 1, \dots, \ell$ ;  
 ii. *restrictive* relative to the data,  
 i.e., no grammar  $\widehat{G}$  such that  $L_{\widehat{G}} \subsetneq L_G$  satisfies the consistency condition (288b.i).

Note that the consistency condition (288b.i) needs to be slightly restated as well, by introducing an existential quantification over the underlying forms and the correspondence relations.

### 6.5.2. Reconstructing faithful underlying forms

A number of scholars in both the computational literature (e.g., Hayes 2004 and Prince and Tesar 2004) and the acquisition literature (e.g., Gnanadesikan 2004 and Pater and Barlow 2003) have proposed that the learner circumvents the lack of underlying forms (and correspondence relations) in the training data (288a.ii) by assuming fully faithful underlying forms, as stated in (289).

- (289) For each training surface form  $\mathbf{b}$  in the training data (288a.ii), the learner assumes the corresponding identity candidate  $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$ , whose existence is guaranteed by the reflexivity axiom (51).

The learner then tackles problem (290) instead of the original problem (288). The crucial difference is that the consistency condition (290a.i) as been restated so as to require the learner to find a grammar which in particular maps each  $\mathbf{b}_i$  to the corresponding identity candidate  $(\mathbf{b}_i, \mathbf{b}_i, \mathbb{I}_{\mathbf{b}_i, \mathbf{b}_i})$ .

- (290) a. *Given*: i. a typology specified through a candidate and a constraint set;  
 ii. aa dataset consisting of  $\ell$  surface forms  $\mathbf{b}_1, \dots, \mathbf{b}_\ell$ .
- b. *Find*: a grammar  $G$  in the typology which is  
 i. *consistent* with the data,  
 i.e.,  $G(\mathbf{b}_i) = (\mathbf{b}_i, \mathbf{b}_i, \mathbb{I}_{\mathbf{b}_i, \mathbf{b}_i})$  for every  $i = 1, \dots, \ell$ ;  
 ii. *restrictive* relative to the data,  
 i.e., no grammar  $\widehat{G}$  such that  $L_{\widehat{G}} \subsetneq L_G$  satisfies the consistency condition (290b.i).

The assumption (289) of faithful underlying forms is *computationally sound* provided each solution of the derived problem (290) also counts as solutions of the original problem (288). This ensures that the learner can solve the original problem (288) by tackling the corresponding derived problem (290) instead. Under which conditions can computational soundness be guaranteed?

### 6.5.3. Idempotency

The consistency condition (288b.i) says that each surface form  $\mathbf{b}_i$  is phonotactically licit relative to the grammar  $G$ , because there exists some underlying form  $\mathbf{a}_i$  which is mapped to  $\mathbf{b}_i$ . The revised consistency condition (290b.i) says that the grammar  $G$  maps each form  $\mathbf{b}_i$  faithfully to itself. The two conditions are equivalent if the grammar  $G$  is idempotent, by the very definition of idempotency provided in section 3.1. If every grammar in the typology explored by the learner is idempotent, the two problems (288) and (290) are thus equivalent: any grammar in the typology solves the former if and only if it solves the latter. And the subroutine (289) is therefore computationally sound: no harm comes by entertaining faithful underlying forms. The sufficient conditions for idempotency developed in chapter 3 thus yield sufficient conditions for the computational soundness of the assumption (289) of faithful underlying forms, leading to the following result.

**Proposition 30.** *Suppose that the learner explores an OT typology corresponding to a constraint set whose faithfulness constraints all satisfy the  $FIC_{comp}$ . For instance, suppose that no correspondence relation in the candidate set breaks any underlying segment and that the faithfulness constraints in the constraint set are drawn from the following list:  $IDENT_\varphi$  (relative to a total feature  $\varphi$ ), segmental MAX and DEP, featural  $MAX_{[\pm\varphi]}$  and  $DEP_{[\pm\varphi]}$  (relative to a possibly partial feature  $\varphi$ ), UNIFORMITY, LINEARITY and ADJACENCY. The assumption (289) of faithful underlying forms is then computationally sound, because the two problems (288) and (290) are equivalent.  $\square$*

Proposition 30 might be useful to investigate specific learnability test cases corresponding to small OT fragments, such as those considered in Hayes (2004) or Prince and Tesar (2004). Yet, the assumption made by the proposition that all the faithfulness constraints in the constraint set satisfy the  $\text{FIC}_{\text{comp}}$  is unfortunately too strong in the general case: it implies that all grammars in the typology explored by the learner are idempotent and it thus contradicts the fact that chain shifts are well attested in both adult and child phonology, as seen in section 6.4.

#### 6.5.4. Benign chain shifts

Fortunately this tension between the existence of chain shifts and the assumption (289) of faithful underlying forms can be relieved. The tension stems from the fact that the assumptions of proposition 30 effectively say that the grammars in the typology explored by the learner are all idempotent. This ensures in turn that the original problem (288) and the derived problem (290) are equivalent: a grammar solves the former if and only if it solves the latter. But the computational soundness of the assumption (289) of faithful underlying forms requires much less: it only requires that a grammar which solves the derived problem (290) also solves the original problem (288)—while the reverse is not needed. The assumptions made by proposition 30 should thus be weakened. We do not need the grammars in the typology to be all idempotent. It is sufficient that for any grammar which fails at idempotency, the typology contains a *companion* grammar which is idempotent and yet makes the same phonotactic distinctions, namely defines exactly the same set of licit surface forms (and therefore also the same set of illicit surface forms, as the two sets are one the complement of the other). Let me thus say that a non-idempotent grammar is *benign* (relative to the typology it belongs to) provided the typology contains a companion grammar  $\hat{G}$  such that  $\hat{G}$  is idempotent and nonetheless  $L_G = L_{\hat{G}}$ . The computational soundness of the assumption (289) of faithful underlying forms is compatible with some grammars in the typology failing at idempotency, as long as the failure is benign. In other words, computational soundness does not require that all grammars in the typology explored by the learner are idempotent and it thus does not exclude chain shifts, as long as all non-idempotent grammars are benign. This observation motivates the search for conditions which guarantee that all grammars in the typology are idempotent or else benign—just as chapter 3 has investigated conditions which guarantee that all grammars in the typology are idempotent.

Unfortunately, at the current stage of the development of computational OT, it does not seem possible to provide general formal sufficient conditions for benignity (mimicking the general formal sufficient condition for idempotency provided by the FIC). To at least start the investigation, I will nonetheless look not at whole grammars but at single chain shifts. A generic chain shift has the shape in (291a). The chain shift is called *benign* provided the typology contains another grammar which enforces the mappings in (291b) and thus “resolves” the shift. Crucially, the set of surface forms which are licit according to the two grammars is the same, namely **b** and **c** at the exclusion of **a**. A chain shift (291a) poses no challenges to the assumption (289) of faithful underlying forms when it is benign, namely when the typology contains the companion grammar (291b).

$$(291) \quad \text{a. } \mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c} \quad \text{b. } \mathbf{a} \rightarrow \mathbf{b} \quad \mathbf{c}$$

The rest of this section thus tackles the following question: under which conditions is a chain shift benign? In other words, under which conditions can I manipulate the ranking corresponding to (291a) and obtain a ranking which corresponds to (291b) instead, thus making sure that both patterns (291) belong to the same OT typology? The gist of the answer that I would like to pursue in future work is the following. As we have seen in chapter 3, in order to flout idempotency and thus derive a chain shift (291a), the faithfulness constraint set needs to contain a faithfulness constraint which fails at the  $\text{FIC}_{\text{comp}}$ . Yet, all *basic* faithfulness constraints satisfy the  $\text{FIC}_{\text{comp}}$  (whereby the common wisdom that chain shifts are problematic for OT). The faithfulness constraints which do fail at the  $\text{FIC}_{\text{comp}}$  are *derived* from *basic*  $\text{FIC}_{\text{comp}}$ -abiding faithfulness constraints, thorough either constraint conjunction or constraint restriction. By simply replacing the derived non- $\text{FIC}_{\text{comp}}$ -complying faithfulness constraint in a ranking (292a) which yields the chain shift (291a) with a corresponding basic  $\text{FIC}_{\text{comp}}$ -complying faithfulness constraint, I hope to obtain a ranking (292b) which yields the phonotactically equivalent grammar (291b).

$$(292) \quad \text{a. } \dots \gg \text{ derived non-}\text{FIC}_{\text{comp}}\text{-complying faithfulness constraint} \gg \dots$$

$$\text{b. } \dots \gg \text{ corresponding basic }\text{FIC}_{\text{comp}}\text{-complying faithfulness constraint} \gg \dots$$

### 6.5.5. Benign chain shifts derived through two markedness constraints

To start investigate benign chain shifts, consider again the concrete chain shift  $s \rightarrow t \rightarrow k$  from subsection 6.4.3. Jesney’s (2007) analysis is repeated in some more detail in (293).

(293)

	$\widehat{M} = *STRID$	$F = IDENT_{[cor]}/\{+str\}$	$M = *TL$
--	------------------------	------------------------------	-----------

/s/			
[s]	*		
$\mathbb{E}$ [t]			*
[k]		*	

	$\widehat{M} = *STRID$	$F = IDENT_{[cor]}/\{+str\}$	$M = *TL$
--	------------------------	------------------------------	-----------

/t/			
[s]	*		
$\mathbb{E}$ [t]			*
$\mathbb{E}$ [k]			

One crucial property of this analysis is that it makes use of two separate markedness constraints to punish [s] and [t], namely  $\widehat{M} = *STRID$  and  $M = *TL$  respectively, with the former ranked above the latter. This allows for further faithfulness constraints to be ranked in between them, such as the (unrestricted) faithfulness constraint  $\widehat{F} = IDENT_{[cor]}$  in (294). This grammar (294) has the same set of phonotactically licit surface forms as the original grammar (293), namely [t] and [k] at the exclusion of [s]. It is furthermore idempotent, because both underlying forms /t/ and /k/ are faithfully mapped to themselves, so that the shift of /t/ to [k] is broken.

(294)

	$\widehat{M} = *STRID$	$F = IDENT_{[cor]}/\{+str\}$	$\widehat{F} = IDENT_{[cor]}$	$M = *TL$
--	------------------------	------------------------------	-------------------------------	-----------

/s/				
[s]	*			
$\mathbb{E}$ [t]				*
[k]		*	*	

	$\widehat{M} = *STRID$	$F = IDENT_{[cor]}/\{+str\}$	$\widehat{F} = IDENT_{[cor]}$	$M = *TL$
--	------------------------	------------------------------	-------------------------------	-----------

/t/				
[s]	*			
$\mathbb{E}$ [t]				*
[k]			*	

The grammar (294) thus shows that the original chain shift (293) is benign.

These considerations hold in complete generality. Thus, consider an arbitrary chain shift  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  which is derived through two markedness constraints  $\widehat{M}$  and  $M$ , which are responsible for punishing  $\mathbf{a}$  and  $\mathbf{b}$ , respectively.  $\widehat{M}$  must be ranked above  $M$ , while the relative ranking of  $\widehat{M}$  and  $F$  does not matter. For concreteness, suppose that  $\widehat{M}$  is also ranked above  $F$ , as in (295). If instead  $\widehat{M}$  is ranked underneath  $F$ , the only difference is that  $F$  can assign no violations to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . The symbol “—” in a cell means that it must have no stars. Empty cells are simply unspecified (not necessarily without stars).

(295)

$\mathbf{a}$	$\widehat{M}$	$F$	$M$
$(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a}, \mathbf{a}})$	*	—	
$\mathbb{E} (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$			*
$(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$		*	

$\mathbf{b}$	$\widehat{M}$	$F$	$M$
$(\mathbf{b}, \mathbf{a}, \mathbb{I}_{\mathbf{b}, \mathbf{a}})$	*		
$(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$		—	*
$\mathbb{E} (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$		—	

Assume that the faithfulness constraint set is *complete*. This means that any candidate which is not the identity candidate violates at least one faithfulness constraint. Completeness ensures that the constraint set contains a faithfulness constraint  $\widehat{F}$  which punishes the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ . Suppose this constraint  $\widehat{F}$  is now ranked between  $\widehat{M}$  and  $M$ , as in (296). No matter the behavior of  $\widehat{F}$  on the other candidates, this grammar (296) has the same set of phonotactically licit surface forms as the original grammar (295), namely  $\mathbf{b}$  and  $\mathbf{c}$  at the exclusion of  $\mathbf{a}$ . It is furthermore idempotent, because both  $\mathbf{b}$  and  $\mathbf{c}$  are faithfully mapped to themselves, so that the shift of  $\mathbf{b}$  to  $\mathbf{c}$  is broken. The grammar (296) thus shows that the original chain shift (295) is benign.

(296)

<b>a</b>	$\widehat{M}$	$F$	$\widehat{F}$	$M$
$(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a},\mathbf{a}})$	*	—		
$\mathbb{E}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$				*
$(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$		*		

<b>b</b>	$\widehat{M}$	$F$	$\widehat{F}$	$M$
$(\mathbf{b}, \mathbf{a}, \mathbb{I}_{\mathbf{b},\mathbf{a}})$	*			
$\mathbb{E}(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$		—		*
$(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$		—	*	

In conclusion, when the constraint set is complete, all chain shifts  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  derived through two separate markedness constraints (one which punishes  $\mathbf{a}$  and possibly  $\mathbf{b}$ ; the other which punishes  $\mathbf{b}$  but not  $\mathbf{a}$ ) are benign. We thus now need to study benignity only for chain shifts  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  derived through a single markedness constraint which is responsible for punishing both  $\mathbf{a}$  and  $\mathbf{b}$ . We consider in turn the three cases where the  $\text{FIC}_{\text{comp}}$  fails because of constraint conjunction, breaking, or constraint restriction.

### 6.5.6. Benign chain shifts derived through constraint conjunction

One strategy to derive chain shifts is to use a conjunction of two faithfulness constraints which both satisfy the  $\text{FIC}_{\text{comp}}$ . I want to investigate whether the chain shift thus derived is benign. As noted in the preceding subsection, it suffices to consider the case where a single markedness constraint is involved in deriving the shift. To start with a concrete case, consider again the chain shift  $\mathbf{a} \rightarrow \mathbf{e} \rightarrow \mathbf{i}$  discussed in subsection 6.4.2. The analysis is repeated in (297), slightly expanded. The conjunction of the two faithfulness constraints  $\text{IDENT}_{[\text{high}]}$  and  $\text{IDENT}_{[\text{low}]}$  does not satisfy the  $\text{FIC}_{\text{comp}}$  and is responsible for the non-idempotent behavior. The single markedness constraint  $\text{RAISE}$  is responsible for penalizing both  $[\mathbf{a}]$  and  $[\mathbf{e}]$  (with the former being penalized more than the latter).

(297)

$/\mathbf{a}/$	$\text{ID}_{[\text{high}]} \wedge \text{ID}_{[\text{low}]}$	$\text{RAISE}$
$[\mathbf{a}]$		**!
$\mathbb{E}[\mathbf{e}]$		*
$[\mathbf{i}]$	*!	

$/\mathbf{e}/$	$\text{ID}_{[\text{high}]} \wedge \text{ID}_{[\text{low}]}$	$\text{RAISE}$
$[\mathbf{a}]$		**!
$[\mathbf{e}]$		*!
$\mathbb{E}[\mathbf{i}]$		

$/\mathbf{i}/$	$\text{ID}_{[\text{high}]} \wedge \text{ID}_{[\text{low}]}$	$\text{RAISE}$
$[\mathbf{a}]$	*!	**
$[\mathbf{e}]$		*!
$\mathbb{E}[\mathbf{i}]$		

Consider now the grammar (298), where it is the constraint  $\text{IDENT}_{[\text{high}]}$  which is ranked above the markedness constraint  $\text{RAISE}$ , while the conjunction  $\text{IDENT}_{[\text{high}]} \wedge \text{IDENT}_{[\text{low}]}$  is low ranked, together with the other conjunct  $\text{IDENT}_{[\text{low}]}$ . This grammar (298) has the same set of phonotactically licit surface forms as the original grammar (297), namely  $[\mathbf{e}]$  and  $[\mathbf{i}]$  at the exclusion of  $[\mathbf{a}]$ . It is furthermore idempotent, because both underlying forms  $/\mathbf{e}/$  and  $/\mathbf{i}/$  are faithfully mapped to themselves, so that the shift of  $/\mathbf{e}/$  to  $[\mathbf{i}]$  is broken.

(298)

$/\mathbf{a}/$	$\text{ID}_{[\text{high}]}$	$\text{RAISE}$
$[\mathbf{a}]$		**!
$\mathbb{E}[\mathbf{e}]$		*
$[\mathbf{i}]$	*!	

$/\mathbf{e}/$	$\text{ID}_{[\text{high}]}$	$\text{RAISE}$
$[\mathbf{a}]$		**!
$\mathbb{E}[\mathbf{e}]$		*
$[\mathbf{i}]$	*!	

$/\mathbf{i}/$	$\text{ID}_{[\text{high}]}$	$\text{RAISE}$
$[\mathbf{a}]$	*!	**
$[\mathbf{e}]$		*!
$\mathbb{E}[\mathbf{i}]$		

The grammar (298) thus shows that the original chain shift (297) is benign.

These considerations hold in complete generality. Thus, consider an arbitrary chain shift  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  which is derived through a single markedness constraint  $M$  together with the conjunction  $F_1 \wedge F_2$  of two faithfulness constraints  $F_1$  and  $F_2$  which both satisfy the  $\text{FIC}_{\text{comp}}$ , as in (299).

(299)

<b>a</b>	$F_1 \wedge F_2$	$M$
$(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a},\mathbf{a}})$	—	**
$\mathbb{E}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$	—	*
$(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$	*	

<b>b</b>	$F_1 \wedge F_2$	$M$
$(\mathbf{b}, \mathbf{a}, \rho_{\mathbf{b},\mathbf{a}})$		**
$(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$	—	*
$\mathbb{E}(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$	—	

Crucially, the conjunction  $F_1 \wedge F_2$  cannot penalize the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  in (299), which would otherwise lose to the identity candidate. This means<sup>5</sup> that one of the two conjoined faithfulness constraints (say,  $F_1$  for concreteness) assigns no violations to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . Since the conjunction  $F_1 \wedge F_2$  assigns violations to the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$ , the conjunct  $F_1$  must in particular assign some violations to that composition candidate. Since  $F_1$  assigns some violations to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$  but not to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , the inequality in the consequent of the  $\text{FIC}_{\text{comp}}$  (265) fails for  $F_1$ . Since  $F_1$  by assumption satisfies the  $\text{FIC}_{\text{comp}}$ , the antecedent of the  $\text{FIC}_{\text{comp}}$  must fail as well for  $F_1$ . In other words,  $F_1$  must assign some violations to the candidate  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$ . Consider now the grammar (300) where it is  $F_1$  which is ranked above the markedness constraint, while the conjunction  $F_1 \wedge F_2$  is low ranked, together with the other conjunct  $F_2$ . This grammar (300) has the same set of phonotactically licit surface forms as the original grammar (299), namely  $\mathbf{b}$  and  $\mathbf{c}$  at the exclusion of  $\mathbf{a}$ . It is furthermore idempotent, because both  $\mathbf{b}$  and  $\mathbf{c}$  are faithfully mapped to themselves, so that the shift of  $\mathbf{b}$  to  $\mathbf{c}$  is broken. The grammar (300) thus shows that the original chain shift (299) is benign.

(300)

<b>a</b>	$F_1$	$M$	<b>b</b>	$F_1$	$M$
$(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a}, \mathbf{a}})$	—	**	$(\mathbf{b}, \mathbf{a}, \rho_{\mathbf{b}, \mathbf{a}})$		**
$\mathbb{E}^{\mathbb{S}}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$	—	*	$\mathbb{E}^{\mathbb{S}}(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$	—	*
$(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$	*		$(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$	*	

In conclusion, all chain shifts  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  derived through a single markedness constraint and a conjunction  $F_1 \wedge F_2$  are benign (modulo footnote 5) whenever the two conjunct faithfulness constraints  $F_1$  and  $F_2$  both satisfy the  $\text{FIC}_{\text{comp}}$ .

### 6.5.7. Benign chain shifts derived through breaking

One strategy to derive a chain shift  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  is to use a plain faithfulness constraint such as DEP or IDENT and to get it to fail at the  $\text{FIC}_{\text{comp}}$  (265) through breaking. I want to investigate whether the chain shift thus derived is benign. If the chain shift involves two markedness constraints which punish  $\mathbf{a}$  and  $\mathbf{b}$  separately, then the chain shift falls within the scope of subsection 6.5.5, which ensures that it is benign whenever the set of faithfulness constraints is complete. Thus assume that there is a unique markedness constraint involved, which is responsible for punishing both  $\mathbf{a}$  and  $\mathbf{b}$ , albeit the former more than the latter, as in (301). Crucially, since the shift involves a single markedness constraint, the faithfulness constraint  $F$  cannot assign any violation to the winning candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ , as it of course assigns no violations to the loser identity candidate  $(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a}, \mathbf{a}})$ .

(301)

<b>a</b>	$F$	$M$	<b>b</b>	$F$	$M$
$(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a}, \mathbf{a}})$	—	**	$(\mathbf{b}, \mathbf{a}, \mathbb{I}_{\mathbf{b}, \mathbf{a}})$		**
$\mathbb{E}^{\mathbb{S}}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$	—	*	$(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b}, \mathbf{b}})$	—	*
$(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a}, \mathbf{b}} \rho_{\mathbf{b}, \mathbf{c}})$	*		$\mathbb{E}^{\mathbb{S}}(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b}, \mathbf{c}})$	—	

Section 3.2 has shown that, in order to derive a chain shift  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$ , we need a faithfulness constraint  $F$  which fails at the  $\text{FIC}_{\text{comp}}$ , repeated once more in (302) for comparison. The tableaux (301) show that, if we furthermore insist on deriving the chain shift through a single markedness constraint being responsible for punishing both  $\mathbf{a}$  and  $\mathbf{b}$ , we need a faithfulness constraint which fails at the implication (303). Note that this implication (303) is weaker than the  $\text{FIC}_{\text{comp}}$ : it is satisfied by any faithfulness constraint which satisfies the  $\text{FIC}_{\text{comp}}$  (but not vice versa<sup>6</sup>). Let me thus refer to this weaker condition as the  $\text{FIC}_{\text{comp}}^{\text{weak}}$ .

<sup>5</sup> Strictly speaking, this entailment only holds if the local domain of conjunction is instantiated only once in the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$ . If that is not the case,  $F_1 \wedge F_2$  could assign no violations to the candidate  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  with  $F_1$  and  $F_2$  both assigning violations to it. In the latter case, the assumption that either  $F_1$  or  $F_2$  assigns no violations to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a}, \mathbf{b}})$  needs to be added as an additional assumption to the reasoning.

<sup>6</sup> Here is a counterexample that shows that the vice versa fails: DEP and IDENT fail at the  $\text{FIC}_{\text{comp}}$  (302) in the presence of breaking, as seen in section 3.3. Yet, they succeed at the weaker  $\text{FIC}_{\text{comp}}^{\text{weak}}$  (303), as guaranteed by the lemma below.

- (302) **If:**  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$   
**Then:**  $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$
- (303) **If:**  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$  and  $F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) = 0$   
**Then:**  $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}) = 0$

Section 3.3 has shown that MAX satisfies the  $FIC_{\text{comp}}$  (and thus also the weaker  $FIC_{\text{comp}}^{\text{weak}}$ ) under no additional assumptions on the correspondence relations, while DEP and IDENT require the additional no-breaking assumption in order to satisfy the  $FIC_{\text{comp}}$ . An analogous reasoning (omitted for brevity<sup>7</sup>) shows that the no-breaking assumption is not needed in order to get DEP and IDENT to satisfy the weaker  $FIC_{\text{comp}}^{\text{weak}}$ , as stated by the following proposition.

**Proposition 31.** *The faithfulness constraints MAX, DEP, and IDENT satisfy the  $FIC_{\text{comp}}^{\text{weak}}$  (303) under no additional assumptions on the correspondence relations.  $\square$*

Let me take stock. In subsection 6.4.4, we have seen that breaking is one of the three possible strategies to derive chain shifts. This subsection has shown that the strategy only works when we use two markedness constraints. The resulting chain shift is therefore benign by virtue of the discussion in subsection 6.5.5.

### 6.5.8. Benign chain shifts derived through constraint restrictions

One final strategy to derive a chain shift  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  is to use a plain faithfulness constraint such as MAX, DEP or IDENT and to get it to fail at the  $FIC_{\text{comp}}$  through restriction. I want to investigate whether the chain shift thus derived is benign. Let me start with the case of  $MAX_R$ . If the chain shift involves two markedness constraints which punish  $\mathbf{a}$  and  $\mathbf{b}$  separately, then the chain shift falls within the scope of subsection 6.5.5, which ensures that it is benign whenever the set of faithfulness constraints is complete. Thus assume that there is a unique markedness constraint involved, which is responsible for punishing both  $\mathbf{a}$  and  $\mathbf{b}$ , albeit the former more than the latter, as in (304).

(304)

$\mathbf{a}$	$MAX_R$	$M$
$(\mathbf{a}, \mathbf{a}, \mathbb{I}_{\mathbf{a},\mathbf{a}})$	—	**
$\mathbb{I}_{\mathfrak{F}}(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$	—	*
$(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$	*	

$\mathbf{b}$	$MAX_R$	$M$
$(\mathbf{b}, \mathbf{a}, \mathbb{I}_{\mathbf{b},\mathbf{a}})$		**
$(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$	—	*
$\mathbb{I}_{\mathfrak{F}}(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$	—	

Crucially,  $MAX_R$  must penalize the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  but neither of the two component candidates  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  and  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ . I consider an arbitrary segment  $\mathbf{a}$  of the string  $\mathbf{a}$  and reason as in (305). In step (305a), I have used the definition of the restricted constraint  $MAX_R$  and the definition of the composition correspondence relation  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ . In step (305b), I have used the fact that  $MAX_R$  does not penalize  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ , which entails that  $\mathbf{a}$  cannot be epenthetic relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . In step (305c), I have noted that the condition  $\forall \mathbf{c}[(\mathbf{b}, \mathbf{c}) \notin \rho_{\mathbf{b},\mathbf{c}}]$  entails that  $\mathbf{b} \notin R(\mathbf{b})$  given the fact that  $MAX_R$  does not penalize  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ .

- (305)  $\mathbf{a}$  violates  $MAX_R$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$
- $\stackrel{(a)}{\Rightarrow} \mathbf{a} \in R(\mathbf{a}), \forall \mathbf{b}[(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \rightarrow \forall \mathbf{c}[(\mathbf{b}, \mathbf{c}) \notin \rho_{\mathbf{b},\mathbf{c}}]]$
- $\stackrel{(b)}{\Rightarrow} \mathbf{a} \in R(\mathbf{a}), \exists \mathbf{b}[(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \text{ and } \forall \mathbf{c}[(\mathbf{b}, \mathbf{c}) \notin \rho_{\mathbf{b},\mathbf{c}}]]$
- $\stackrel{(c)}{\Rightarrow} \mathbf{a} \in R(\mathbf{a}), \exists \mathbf{b}[(\mathbf{a}, \mathbf{b}) \in \rho_{\mathbf{a},\mathbf{b}} \text{ and } \forall \mathbf{c}[(\mathbf{b}, \mathbf{c}) \notin \rho_{\mathbf{b},\mathbf{c}}] \text{ and } \mathbf{b} \notin R(\mathbf{b})]$

This chain of implications says in particular that, since the restricted constraint  $MAX_R$  penalizes the composition candidate  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ , the corresponding unrestricted constraint MAX also penalizes  $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ , because there exists some segment  $\mathbf{b}$  which does not satisfy the restriction  $R$  and has no correspondent

<sup>7</sup> The intuition is straightforward. The proofs of the fact that MAX, DEP, and IDENT satisfy the  $FIC_{\text{comp}}$  (302) presented in appendices A.1, A.2, and A.3 use the antecedent  $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$  of the  $FIC_{\text{comp}}$  to show that every violation of  $F$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  corresponds to a violation of  $F$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ . The no-breaking assumption was then used to establish the injectivity of this correspondence. But the additional assumption  $F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) = 0$  in the antecedent of the  $FIC_{\text{comp}}^{\text{weak}}$  (303) says that there are no violations of  $F$  relative to  $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$  to start with. Hence, there can be no violations of  $F$  relative to  $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$  either. This establishes the consequent in the  $FIC_{\text{comp}}^{\text{weak}}$  (303).



in **c**. Consider now the grammar (306) where it is the unrestricted MAX which is ranked above the markedness constraint, while the restricted counterpart  $\text{MAX}_R$  is low ranked. This grammar (306) has the same set of phonotactically licit surface forms as the original grammar (304), namely **b** and **c** at the exclusion of **a**. It is furthermore idempotent, because both **b** and **c** are faithfully mapped to themselves, so that the shift of **b** to **c** is broken. The grammar (306) thus shows that the original chain shift (304) is benign.

(306)

<b>a</b>	MAX	M
( <b>a</b> , <b>a</b> , $\mathbb{I}_{\mathbf{a},\mathbf{a}}$ )	—	**
$\mathbb{E}^S$ ( <b>a</b> , <b>b</b> , $\rho_{\mathbf{a},\mathbf{b}}$ )	—	*
( <b>a</b> , <b>c</b> , $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ )	*	

<b>b</b>	MAX	M
( <b>b</b> , <b>a</b> , $\mathbb{I}_{\mathbf{b},\mathbf{a}}$ )		**
$\mathbb{E}^S$ ( <b>b</b> , <b>b</b> , $\mathbb{I}_{\mathbf{b},\mathbf{b}}$ )	—	*
( <b>b</b> , <b>c</b> , $\rho_{\mathbf{b},\mathbf{c}}$ )	*	

In conclusion, all chain shifts  $\mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{c}$  derived through a single markedness constraint and a restricted constraint  $\text{MAX}_R$  are benign. Analogous considerations hold for  $\text{IDENT}_R$ . This reasoning instead fails for  $\text{DEP}^S$ . By repeating the reasoning in (305) for DEP rather than MAX, the fact that (**a**, **c**,  $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ ) violates  $\text{DEP}^S$  entails that (**a**, **b**,  $\rho_{\mathbf{a},\mathbf{b}}$ ) violates DEP, not that (**b**, **c**,  $\rho_{\mathbf{b},\mathbf{c}}$ ) violates DEP.

### 6.5.9. Summary

Idempotency is a property of the phonology, not of the phonotactics. A certain phonotactic pattern might correspond to various and very different phonological grammars, some of which are idempotent while some others are not. The computational soundness of the assumption (289) of faithful underlying forms only requires idempotency to hold for one of the grammars which yield the target phonotactic pattern, not for all of them. The assumption (289) of faithful underlying forms is thus not a priori incompatible with the typology containing chain shifts. This section has sketched a way to capitalize on this intuition.

## A. Proofs

### A.1. Proof of proposition 27

Throughout this appendix, let me denote by (307) the fact that the surface form **b** can be obtained from the underlying form **a** through the SPE *derivation* which applies rule  $\mathbb{R}_1$  first, immediately followed by rule  $\mathbb{R}_2$ , and so. The mode of rule application (whether each rule applies simultaneously to all targets or sequentially, say, from left to right) is not crucial for what follows.

$$(307) \quad \langle \mathbf{a}, \mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_n, \dots \rangle = \mathbf{b}$$

A rule  $\mathbb{R}_n$  is said to apply *vacuously* in the derivation  $\langle \mathbf{a}, \mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{n-1}, \mathbb{R}_n, \mathbb{R}_{n+1}, \dots \rangle$  provided condition (308) holds: at the stage in the derivation when rule  $\mathbb{R}_n$  is applied, it has no effect.<sup>8</sup>

$$(308) \quad \langle \mathbf{a}, \mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{n-1}, \mathbb{R}_n \rangle \neq \langle \mathbf{a}, \mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{n-1} \rangle.$$

With this notation, rule  $\mathbb{B}$  *counter-feeds* rule  $\mathbb{A}$  relative to the derivation  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$  provided the two conditions (309) hold. The first condition (309a) says that the rule  $\mathbb{A}$  applies vacuously in the original derivation  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$ . The second condition (232b) says that the rule  $\mathbb{A}$  instead applies non-vacuously in the swapped derivation  $\langle \mathbf{a}, \mathbb{B}, \mathbb{A} \rangle$ .

$$(309) \quad \begin{array}{l} \text{a. } \langle \mathbf{a}, \mathbb{A} \rangle = \mathbf{a} \\ \text{b. } \langle \mathbf{a}, \mathbb{B}, \mathbb{A} \rangle \neq \langle \mathbf{a}, \mathbb{B} \rangle \end{array}$$

<sup>8</sup> According to condition (308), vacuity of the rule  $\mathbb{R}_n$  only depends on the underlying form **a** and on the rules  $\mathbb{R}_1, \dots, \mathbb{R}_{n-1}$  which precede rule  $\mathbb{R}_n$ , not on the following rules  $\mathbb{R}_{n+1}, \dots$  which follow rule  $\mathbb{R}_n$ . Condition (i) provides a natural alternative to condition (308), which makes vacuity also possibly dependent on the later rules. According to this alternative condition (i), rule  $\mathbb{R}_n$  is vacuous in the derivation provided it can be omitted without changing the outcome of the derivation.

$$(i) \quad \langle \mathbf{a}, \mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{n-1}, \mathbb{R}_n, \mathbb{R}_{n+1}, \dots \rangle = \langle \mathbf{a}, \mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{n-1}, \mathbb{R}_{n+1}, \dots \rangle.$$

Conditions (308) and (i) coincide for the last rule in a derivation. For any other rule, condition (308) entails condition (i) but the vice versa fails, as shown by Duke-of-York derivations. The choice between these two definitions of rule vacuity is irrelevant for what follows.

Proposition 27 from subsection 6.2.1 can be made explicit as follows, where the expression “under mild assumptions” used in the informal formulation of the proposition provided in subsection 6.2.1 has been made explicit through the (underlined> condition that the derivation  $\langle\langle\mathbf{a}, \mathbb{B}\rangle, \mathbb{A}, \mathbb{B}\rangle$  is not a Duke-of-York. The latter assumption is mild because Duke-of-Yorks are uncommon.

**PROPOSITION 27** *Suppose that  $\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle$  is a counter-feeding derivation and that the derivation  $\langle\langle\mathbf{a}, \mathbb{B}\rangle, \mathbb{A}, \mathbb{B}\rangle$  is not a Duke-of-York, namely that it is not the case that the two following conditions (310) both hold.*

- (310) a.  $\langle\mathbf{a}, \mathbb{B}\rangle \neq \langle\langle\mathbf{a}, \mathbb{B}\rangle, \mathbb{A}\rangle$  (first Duke-of-York condition)  
 b.  $\langle\mathbf{a}, \mathbb{B}\rangle = \langle\langle\mathbf{a}, \mathbb{B}\rangle, \mathbb{A}, \mathbb{B}\rangle$  (second Duke-of-York condition)

The SPE phonology corresponding to the ordered rules  $\langle\mathbb{A}, \mathbb{B}\rangle$  then yields a chain shift:  $\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle = \mathbf{b}$  but  $\langle\mathbf{b}, \mathbb{A}, \mathbb{B}\rangle \neq \mathbf{b}$ .  $\square$

PROOF. Condition (309b) in the definition of counter-feeding says that the first Duke-of-York condition (310a) holds. The second Duke-of-York condition (310b) therefore fails. Let  $\mathbf{b}$  the result of applying the SPE phonology  $\langle\mathbb{A}, \mathbb{B}\rangle$  to the underlying form  $\mathbf{a}$ , namely  $\mathbf{b} = \langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle$ . The claim follows by applying the SPE phonology  $\langle\mathbb{A}, \mathbb{B}\rangle$  to  $\mathbf{b}$  as in (311).

$$(311) \quad \langle\mathbf{b}, \mathbb{A}, \mathbb{B}\rangle \stackrel{(a)}{=} \langle\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle, \mathbb{A}, \mathbb{B}\rangle \stackrel{(b)}{=} \langle\langle\mathbf{a}, \mathbb{B}\rangle, \mathbb{A}, \mathbb{B}\rangle \neq \langle\mathbf{a}, \mathbb{B}\rangle \stackrel{(d)}{=} \langle\langle\mathbf{a}, \mathbb{A}\rangle, \mathbb{B}\rangle \stackrel{(e)}{=} \mathbf{b}$$

In step (311a), I have used the definition of the form  $\mathbf{b}$  as the result of applying  $\langle\mathbb{A}, \mathbb{B}\rangle$  to  $\mathbf{a}$ . In step (311b), I have used the first condition (309a) of the definition of counter-feeding, which says that  $\mathbb{A}$  applies vacuously to  $\mathbf{a}$ . In step (311c), I have used the fact that the second Duke-of-York condition (310b) fails. In step (311d), I have used again the first condition (309a) of the definition of counter-feeding. Finally in step (311e), I have used again the definition of  $\mathbf{b}$ .  $\square$

The assumption that  $\langle\langle\mathbf{a}, \mathbb{B}\rangle, \mathbb{A}, \mathbb{B}\rangle$  is not a Duke-of-York derivation is crucial in order for proposition 27 to hold. Here is a counterexample. Nootka (also known as Nuuchahnulth; Sapir and Swadesh 1978; McCarthy 1999; Kenstowicz and Kisseberth 1977) has the two rules (312). The labialization rule  $\mathbb{A}$  is illustrated by the mapping  $/\text{haju}+\text{qi}/ \rightarrow [\text{hajuq}^w\text{i}]$  (‘ten on top’). The delabialization rule  $\mathbb{B}$  is illustrated by the mapping  $/+\text{a:k}^w+\text{fit}\text{t}/ \rightarrow [+a:k\text{fit}\text{t}]$  (‘to take pity on’).

- (312)  $\mathbb{A}$ :  $[+\text{dor}] \rightarrow [+r\text{nd}] / [+r\text{nd}] \_\_\_$  (Dorsals labialization after round vowels)  
 $\mathbb{B}$ :  $[+\text{dor}] \rightarrow [-r\text{nd}] / \_\_\_ ]_\sigma$  (Delabialization of dorsal codas)

Let  $\mathbf{a} = \text{uq}^w$  and  $\mathbf{b} = \text{uq}$ . The derivation  $\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle = \mathbf{b}$  is an instance of counter-feeding, because the labialization rule  $\mathbb{A}$  applies vacuously in that derivation whereas it applies non vacuously in the swapped derivation  $\langle\mathbf{a}, \mathbb{B}, \mathbb{A}\rangle$ . Yet, no chain shift arises in this case, because  $\langle\mathbf{b}, \mathbb{A}, \mathbb{B}\rangle = \mathbf{b}$ . And indeed in this case both Duke-of-York conditions (310a) and (310b) hold, so that proposition 27 does not apply.

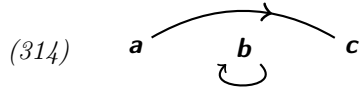
## A.2. Proof of proposition 28

With the notation introduced in appendix A.1, rule  $\mathbb{B}$  *counter-bleeds* rule  $\mathbb{A}$  relative to the derivation  $\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle$  provided the two conditions (313) hold. The first condition (313a) says that the rule  $\mathbb{A}$  applies non-vacuously in the original derivation  $\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle$ . The second condition (235b) says that the rule  $\mathbb{A}$  instead applies vacuously in the swapped derivation  $\langle\mathbf{a}, \mathbb{B}, \mathbb{A}\rangle$ .

- (313) a.  $\langle\mathbf{a}, \mathbb{A}\rangle \neq \mathbf{a}$   
 b.  $\langle\mathbf{a}, \mathbb{B}, \mathbb{A}\rangle = \langle\mathbf{a}, \mathbb{B}\rangle$

Proposition 28 from subsection 6.2.2 can be made explicit as follows, where the expression “under mild assumptions” used in the informal formulation of the proposition provided in subsection 6.2.2 has been made explicit through the (underlined> condition that  $\mathbb{A}$  does not bleed  $\mathbb{B}$  in the derivation  $\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle$ . The latter assumption is mild because its failure would mean that the derivation  $\langle\mathbf{a}, \mathbb{B}, \mathbb{A}\rangle$  is a case of *mutual bleeding* (Baković 2011), which is uncommon.

**PROPOSITION 28** *Let  $\mathbf{b}$  the result of applying rule  $\mathbb{B}$  to an underlying form  $\mathbf{a}$ , namely  $\mathbf{b} = \langle\mathbf{a}, \mathbb{B}\rangle$ . Let  $\mathbf{c}$  be the result of applying rule  $\mathbb{A}$  first followed by rule  $\mathbb{B}$  to that same underlying form  $\mathbf{a}$ , namely  $\mathbf{c} = \langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle$ . If the latter derivation is a counter-bleeding derivation, then  $\langle\mathbf{b}, \mathbb{A}, \mathbb{B}\rangle = \mathbf{b}$ . If  $\mathbb{A}$  does not bleed  $\mathbb{B}$  in the derivation  $\langle\mathbf{a}, \mathbb{A}, \mathbb{B}\rangle$ , the resulting pattern*



is a saltation because **a** is closer to **b** than to **c**, as **b** is obtained from **a** through rule  $\mathbb{B}$  only while **c** is obtained from **a** by through both rules  $\mathbb{A}$  and  $\mathbb{B}$  which are both non-vacuous.  $\square$

PROOF. The form **b** is faithfully mapped to itself by the SPE phonology  $\langle \mathbb{A}, \mathbb{B} \rangle$ , as shown in (315). In step (315a), I have used the definition of the form **b** as the result of applying rule  $\mathbb{A}$  to **a**. In step (315b), I have used the second condition (313b) of the definition of counter-bleeding. In step (315c), I have used the fact that SPE rules apply exhaustively. Finally in step (315d), I have used again the fact that **b** is the result of applying rule  $\mathbb{B}$  to **a**.

$$(315) \quad \langle \mathbf{b}, \mathbb{A}, \mathbb{B} \rangle \stackrel{(a)}{=} \langle \langle \mathbf{a}, \mathbb{B} \rangle, \mathbb{A}, \mathbb{B} \rangle = \langle \langle \mathbf{a}, \mathbb{B}, \mathbb{A} \rangle, \mathbb{B} \rangle \stackrel{(b)}{=} \langle \langle \mathbf{a}, \mathbb{B} \rangle, \mathbb{B} \rangle \stackrel{(c)}{=} \langle \mathbf{a}, \mathbb{B} \rangle \stackrel{(d)}{=} \mathbf{b}$$

By assumption, it is not the case that  $\mathbb{A}$  bleeds  $\mathbb{B}$  relative to the derivation  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$ . In other words, both rules  $\mathbb{A}$  and  $\mathbb{B}$  apply non-vacuously in that derivation. The pattern (314) therefore counts as a saltation because **c** =  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$  is separated from **a** through two non-vacuous rules while **b** =  $\langle \mathbf{a}, \mathbb{B} \rangle$  is separated from **a** through only one of those two rules.  $\square$

The assumption that  $\mathbb{A}$  does not bleed  $\mathbb{B}$  in  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$  is crucial in order for proposition 28 to hold. Here is a counterexample (based on Baković 2011). Lardil (Hale 1973; Hayes 2009) has the two rules in (316). The glide-epenthesis rule  $\mathbb{A}$  is illustrated by the mapping  $/\text{papi}+\text{u}\downarrow/ \rightarrow [\text{papiwu}\downarrow]$  ('father's mother-ACC-FUT'). The vowel deletion rule  $\mathbb{B}$  is illustrated by the mapping  $/\text{t}^{\text{j}}\text{æmpæ}+\text{u}\downarrow/ \rightarrow [\text{t}^{\text{j}}\text{æmpæ}\downarrow]$  ('mother's father-ACC-FUT').

$$(316) \quad \begin{array}{ll} \mathbb{A}: \emptyset \rightarrow \text{w} / \text{i} \_ \_ \text{u} & \text{(Glide epenthesis)} \\ \mathbb{B}: \text{V} \rightarrow \emptyset / \text{V} \_ \_ & \text{(Vowel deletion)} \end{array}$$

Let **a** = *iu*, **b** = *i* and **c** = *iwu*. Hence **b** =  $\langle \mathbf{a}, \mathbb{B} \rangle$  and the derivation **c** =  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$  is a case of counter-bleeding. Yet, no saltation arises in this case: there is no sense in which the surface form **c** = *iwu* is further away than the surface form **b** = *i* from the underlying form **a** = *iu*. And indeed rule  $\mathbb{A}$  does bleed rule  $\mathbb{B}$  in the derivation  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$ , so that proposition 28 does not apply.

The implication from counter-bleeding to saltations established by proposition 28 is not invertible: saltations can arise without counter-bleeding, as shown by the following counterexample (from Baković 2011). Turkish (Kenstowicz and Kisseberth 1979) has the two rules (317). The *s*/*j*-deletion rule  $\mathbb{A}$  is illustrated by the mapping  $/\text{t}^{\text{j}}\text{an}+\text{su}\downarrow/ \rightarrow [\text{t}^{\text{j}}\text{an}\downarrow]$ . The *g*-deletion rule  $\mathbb{B}$  is illustrated by the mapping  $/\text{bebeg}+\text{i}\downarrow/ \rightarrow [\text{bebei}\downarrow]$ .

$$(317) \quad \begin{array}{ll} \mathbb{A}: \text{s/j} \rightarrow \emptyset / \text{C} \_ \_ & \text{(post-consonantal s/j-deletion)} \\ \mathbb{B}: \text{g} \rightarrow \emptyset / \text{V} \_ \_ \text{V} & \text{(intervocalic g-deletion)} \end{array}$$

Let **a** =  $/\text{VgsV}/$ , **b** =  $/\text{VsV}/$ , and **c** =  $[\text{VV}]$ . Hence **b** =  $\langle \mathbf{a}, \mathbb{B} \rangle$ . The derivation **c** =  $\langle \mathbf{a}, \mathbb{A}, \mathbb{B} \rangle$  is not an instance of counter-bleeding (Baković 2011 dubs such cases *self-destructive feeding on the environment*). Yet, the corresponding phonology yields a saltation: the underlying form **a** =  $/\text{VgsV}/$  is mapped to the surface form **c** =  $[\text{VV}]$  not to the closer and phonotactically licit surface form **b** =  $/\text{VsV}/$ .

## Bibliography

- Alderete, John. 2001. *Morphologically governed accent in Optimality Theory*. Outstanding Dissertations in Linguistics. New York.: Routledge.
- Alderete, John. 2008. Using learnability as a filter on factorial typology: A new approach to Anderson and Browne's generalization. *Lingua* 118:1177–1220.
- Baković, Eric. 2011. Opacity and ordering. In *The handbook of phonological theory*, ed. John A. Goldsmith, Jason Riggle, and Alan C. L. Yu, 40–67. Wiley Blackwell.
- Bane, Max, Jason Riggle, and Morgan Sonderegger. 2010. The VC dimension of constraint-based grammars. *Lingua* 120.5:1194–1208.
- Barlow, Jessica A. 2001. The structure of /s/-sequences: evidence from a disordered system. *Journal of child Language* 28:291–324.
- Barner, David, and Jesse Snedeker. 2005. Quantity judgments and individuation: evidence that mass nouns count. *Cognition* 97:41–66.
- Barton, G. Edward, Robert Berwick, and Eric Sven Ristad. 1987. *Computational complexity and natural language*. Cambridge, MA: MIT Press.
- Beckman, Mary E. 1997. Positional faithfulness, positional neutralization and shona vowel harmony. *Phonology* 14:1–46.
- Beckman, Mary E. 1999. *Positional faithfulness: An Optimality Theoretic treatment of phonological asymmetries*. New York: Garland Publishing (Outstanding dissertations in Linguistics).
- Benua, Laura. 2000. *Phonological relations between words*. New York: Garland.
- Bermúdez-Otero, Ricardo. 2007. Morphological structure and phonological domains in Spanish denominal derivation. In *Optimality-theoretic studies in Spanish phonology*, ed. Fernando Martínez-Gil and Sonia Colina, 278–311. Amsterdam: John Benjamins.
- Bernhardt, Barbara Handford, and Joseph Paul Stemberger. 1998. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. Academic Press.
- Bertsimas, Dimitris, and John N. Tsitsiklis. 1997. *Linear optimization*. Athena Scientific.
- Berwick, Robert. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.
- Blaho, Sylvia, Patrik Bye, and Martin Krämer. 2007. *Freedom of analysis?*. Walter de Gruyter.
- Boersma, Paul. 1997. How we learn variation, optionality and probability. In *Proceedings of the Institute of Phonetic Sciences (IFA) 21*, ed. Rob van Son, 43–58. University of Amsterdam: Institute of Phonetic Sciences.
- Boersma, Paul. 1998. Functional phonology. Doctoral Dissertation, University of Amsterdam, The Netherlands. The Hague: Holland Academic Graphics.
- Boersma, Paul, and Bruce Hayes. 2001. Empirical tests for the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45–86.
- Boersma, Paul, and Joe Pater. 2007. Convergence properties of a gradual learner for harmonic grammar. In *Proceedings of NELS 38*.
- Boersma, Paul, and Joe Pater. to appear. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In *Harmonic Grammar and Harmonic Serialism*, ed. John McCarthy and Joe Pater. London: Equinox Press.
- Bolognesi, Roberto. 1998. The phonology of Campidanian Sardinian: A unitary account of self-organizing structure. Doctoral Dissertation, University of Amsterdam, Amsterdam.
- Buccola, Brian. 2013. On the expressivity of Optimality Theory versus ordered rewrite rules. In *Proceedings of Formal Grammar 2012 and 2013*, ed. Glyn Morrill and Mark-Jan Nederhof, Lecture Notes in Computer Science. Springer.
- Burzio, Luigi. 1998. Multiple correspondence. *Lingua* 104:79–109.

- Cardinaletti, Anna, and Giuliana Giusti. 2006. The syntax of quantified phrases and quantitative clitics. In *The blackwell companion to syntax, volume v*, ed. Martin Everaert and Henk van Riemsdijk, 23–93. Blackwell Publishing.
- Carlson, Gregory N. 1977. Reference to kinds in English. Doctoral Dissertation, University of Massachusetts at Amherst. Published in 1980 by Garland Press, New York.
- Carpenter, Angela. 2002. Noncontiguous metathesis and ADJACENCY. In *Papers in Optimality Theory*, ed. Angela Carpenter, Andries Coetzee, and Paul de Lacy, volume 2, 1–26. Amherst, MA: GLSA.
- Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73:493–533.
- Casali, Roderic F. 1998. *Resolving hiatus*. Outstanding dissertations in Linguistics. New York: Garland.
- Chierchia, Gennaro. 1995. Individual-level predicates as inherent generics. In *The Generic Book*, ed. Gregory N. Carlson and Francis Jeffry Pelletier, 125–175. The Univ. of Chicago Press.
- Chierchia, Gennaro. 1996. Plurality of mass nouns and the notion of ‘semantic parameter’. In *Events and grammar*, ed. S. Rothstein. Kluwer Academic Publisher.
- Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In *Structures and beyond*, ed. Adriana Belletti. Oxford: Oxford University Press.
- Chierchia, Gennaro, Danny Fox, and Benjamin Spector. 2012. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In *Handbook of Semantics*, ed. Paul Portner, Claudia Maienborn, and Klaus von Stechow, volume 3. Berlin, New York: Mouton de Gruyter.
- Cho, Mi-Hui, and Shinsook Lee. 2000. Ordering problems in phonological acquisition. *Studies in Phonetics, Phonology and Morphology* 6:137–157.
- Cho, Mi-Hui, and Shinsook Lee. 2003. The acquisition of fricatives: chain shift cases of english and korean. *Studies in Phonetics, Phonology and Morphology* 9:485–498.
- Clark, Alexander, and Shalom Lappin. 2011. *Linguistic nativism and the poverty of the stimulus*. Wiley-Blackwell.
- Coetzee, Andries W., and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31:47–89.
- Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26:289–337.
- Coetzee, Andries W., and Joe Pater. 2011. The place of variation in phonological theory. In *Handbook of phonological theory*, ed. John Goldsmith, Jason Riggle, and Alan Yu, 401–434. Cambridge: Blackwell.
- Crowhurst, Megan, and Mark Hewitt. 1997. Boolean operations and constraint interactions in Optimality Theory. Ms. University of North Carolina and Brandeis University. Available as ROA-229.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, Mass.: MIT Press.
- Dinnsen, Daniel A., and Jessica A. Barlow. 1998. On the characterization of a chain shift in normal and delayed phonological acquisition. *Journal of Child Language* 25:61–94.
- Dinnsen, Daniel A., Christopher R. Green, Judith A. Gierut, and Michelle L. Morrisette. 2011. On the anatomy of a chain shift. *Journal of Linguistics* 47.2:1–25.
- Dinnsen, Daniel A., Kathleen M. O’Connor, and Judith A. Gierut. 2001. The puzzle-puddle-pickle problem and the duke-of-york gambit in acquisition. *Journal of Linguistics* 37:503–525.
- Downing, Laura J. 1998. On the prosodic misalignment of onsetless syllables. *Natural Language and Linguistic Theory* 16:1–52.
- Downing, Laura J. 2000. Morphological and prosodic constraints on Kinande verbal reduplication. *Phonology* 17:1–38.
- Elbourne, Paul. 1999. Some correlations between semantic plurality and quantifier scope. In *NELS 29. Proceedings of the North East Linguistic Society*, ed. Pius Tamanji, Masako Hirotoni, , and Nancy Hall, 81–92. Amherst: GLSA.
- Fernald, Theodore B. 2000. *Predicates and temporal arguments*. New York, Oxford: Oxford University Press.
- Ferreira, Marcelo. 2005. Event quantification and plurality. Doctoral Dissertation, MIT.
- Fikkert, Paula. 1994. On the acquisition of prosodic structure. Doctoral Dissertation, University of Leiden. Published by The Hague: Holland Academic Graphics.
- Flack, Kathryn. 2007. Inducing functionally grounded constraints. In *Papers in theoretical and computational phonology*, ed. Michael Becker, UMOP (University of Massachusetts Occasional Papers) 36, 13–44. Amherst, MA: GLSA.

- Fleischhacker, Heidi Anne. 2001. Cluster-dependent epenthesis asymmetries. In *UCLA working papers in linguistics no. 7 (papers in phonology 5)*, ed. Adam Albright and Taehong Cho, 71–116.
- Fleischhacker, Heidi Anne. 2005. Similarity in phonology: Evidence from reduplication and loan adaptation. Doctoral Dissertation, University of California, Los Angeles.
- Fodor, Janet Dean, and William Gregory Sakas. 2005. The Subset Principle in syntax: costs of compliance. *Linguistics* 41:513–569.
- Fox, Danny. 1995. Economy and scope. *Natural Language Semantics* 3:283–341.
- Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.
- Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics*, ed. Uli Sauerland and Penka Stateva, 71–120. New York: Palgrave Macmillan.
- Fox, Danny, and Roni Katzir. 2009. Alternatives for implicature and focus. In *Presuppositions and implicatures. proceedings of the mit-paris workshop*, ed. Paul Égré and Giorgio Magri, 101–110. Cambridge, MA: MIT Working Papers in Linguistics, volume 60.
- Fox, Danny, and Benjamin Spector. to appear. Economy and embedded exhaustification. *Natural Language Semantics* .
- Geurts, Bart. 2010. *Quantity implicatures*. Cambridge University Press.
- Gillon, B., E. Kehayia, and V. Taler. 1999. The mass/count distinction: Evidence from on-line psycholinguistic performance. *Brain and Language* 68:205–211.
- Gnanadesikan, Amalia E. 1997. Phonology with ternary scales. University of Massachusetts Amherst.
- Gnanadesikan, Amalia E. 2004. Markedness and faithfulness constraints in child phonology. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 73–108. Cambridge: Cambridge University Press. Circulated since 1995.
- Goldrick, Matthew A. 2001. Turbid output representations and the unity of opacity. In *Proceedings of the Northeast Linguistics Society (NELS) 30*, ed. M. Hirotani, A. Coetzee, N. Hall, and J.-Y. Kim, volume 1, 231–245. Amherst, MA: GLSA.
- Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. *Current trends in linguistics* 11:401–458.
- Hall, Nancy. 2011. Vowel epenthesis. In *The blackwell companion to phonology*, ed. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 1576–1596. John Wiley & Sons.
- Harris, John. 2011. Deletion. In *The blackwell companion to phonology*, ed. Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, volume 3, 1597–1621. Oxford: Wiley-Blackwell.
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 158–203. Cambridge: Cambridge University Press.
- Hayes, Bruce. 2009. *Introductory phonology*. Wiley-Blackwell.
- Hayes, Bruce, and James White. to appear. Saltation and the p-map. *Phonology* .
- Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.
- Heim, Irene. 1991. Artikel und definitheit. In *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, ed. Arnim von Stechow and Dieter Wunderlich, 487–535. De Gruyter: Berlin.
- Heinz, Jeffrey. 2005. Reconsidering linearity: Evidence from CV metathesis. In *Proceedings of WCCFL 24*, ed. John Alderete, Chung-hye Han, and Alexei Kochetov, 200–208. Somerville, MA, USA: Cascadia Press.
- Heinz, Jeffrey. 2011. Computational phonology. part i: Foundations. *Linguistic Compass* 5.4:140–152.
- Horn, Laurence R. 2005. The border wars: a neo-gricean perspective. In *Where Semantics meets Pragmatics*, ed. Ken Turner and Klaus von Heusinger. Elsevir.
- Hualde, José I. 1989. Autosegmental and metrical spreading in the vowel harmony systems of northwestern Spain. *Linguistics* 27:773–805.
- Hume, Elizabeth. 1998. Metathesis in phonological theory: The case of Leti. *Lingua* 104:147–186.
- Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7:217–259.
- Jäger, Gerhard. 2001. Topic-comment structure and the contrast between stage level and individual level predicates. *Journal of Semantics* 18:83–126.
- Jarosz, Gaja. 2013. Learning with hidden structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretative Parsing. *Phonology* 30:27–71.

- Jesney, Karen. 2005. Chain shifts in phonological acquisition. Doctoral Dissertation, University of Calgary. Master Thesis.
- Jesney, Karen. 2007. Child chain shifts as faithfulness to input prominence. In *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition*, 188–199. Somerville, MA: Cascadilla Proceedings Project.
- Jesney, Karen, and Anne-Michelle Tessier. 2011. Biases in Harmonic Grammar: the road to restrictive learning. *Natural Language and Linguistic Theory* 29:251–290.
- Jusczyk, P. W., A. D. Friederici, J. M. I. Wessels, V. Y. Svenkerud, and A. Jusczyk. 1993. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32:402–420.
- Katzir, Roni, and Raj Singh. 2014. Economy of structure and information. In *Proceedings of Sinn und Bedeutung 2014*, –.
- Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82:536–574.
- Keller, Frank. 2000. Gradience in grammar. Experimental and computational aspects of degrees of grammaticality. Doctoral Dissertation, University of Edinburgh, England.
- Kenstowicz, Michael, and Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.
- Kenstowicz, Michael, and Charles W. Kisseberth. 1979. *Generative phonology: Description and theory*. New York: Academic Press.
- Kirchner, Robert. 1995. Going the distance: Synchronic chain shifts in Optimality Theory. Ms. UCLA. Available as ROA-66.
- Kirchner, Robert. 1996. Synchronic chain-shifts in Optimality Theory. *LI* 27.2:341–350.
- Kratzer, Angelika. 1995. Stage-level and individual-level predicates. In *The Generic Book*, ed. Gregory N. Carlson and Francis Jeffrey Pelletier, 125–175. The Univ. of Chicago Press.
- Kratzer, Angelika. in progress. The event argument and the semantics of verbs. Univ. of Massachusetts at Amherst; available online at [www.semanticarchive.net](http://www.semanticarchive.net).
- Kubozono, Haruo, Junko Ito, and Armin Mester. 2008. Consonant gemination in Japanese loanword phonology. In *18th International Congress of Linguistics*. Seoul.
- de Lacy, Paul. 2006. *Markedness. reduction and preservation in phonology*. Cambridge University Press.
- Landman, Fred. 1989. Groups i. *Linguistics and Philosophy* 12:559–605.
- Landman, Fred. 1991. *Structures for semantics*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Lauer, Sven. 2014. Mandatory implicatures in gricean pragmatics. In *Proceedings of "Formal and Experimental Pragmatics 2014"*, ed. J. Degen, M. Franke, and N. Goodman, 21–28.
- Lee, Shin Sook. 2000. Chain shift in second language phonological acquisition. *Journal of the Pan-Pacific Association of Applied Linguistics* 4:175–199.
- Legendre, G eraldine, Yoshiro Miyata, and Paul Smolensky. 1990a. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: An application. In *Annual conference of the Cognitive Science Society 12*, ed. Morton Ann Gernsbacher and Sharon J. Derry, 884–891. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Legendre, G eraldine, Yoshiro Miyata, and Paul Smolensky. 1990b. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Annual conference of the Cognitive Science Society 12*, ed. Morton Ann Gernsbacher and Sharon J. Derry, 388–395. Mahwah, NJ: Lawrence Erlbaum.
- Legendre, G eraldine, Antonella Sorace, and Paul Smolensky. 2006. The optimality theory/harmonic grammar connection. In *The harmonic mind*, ed. Paul Smolensky and G eraldine Legendre, 903–966. Cambridge, MA: MIT Press.
- Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Meaning, use and interpretation in language*, ed. R. Bauerle, C. Schwartze, and A. von Stechow, 302–323. Berlin: Mouton de Gruyter.
- Link, Godehard. 1984. Hydras. on the logic of relative clause constructions with multiple heads. In *Varieties of formal semantics*, ed. F. Landman and F. Veltman. Dordrecht: GRASS 3, Foris.
- Lombardi, Linda. 2001. Why place and voice are different: Constraint interactions and feature faithfulness in optimality theory. In *Segmental phonology in Optimality Theory: Constraints and representations*, ed. Linda Lombardi, 13–45. Cambridge University Press.

- Lubowicz, Anna. 2002. Derived environment effects in Oprimality Theory. *Lingua* 112:243–280.
- Lubowicz, Anna. 2011. Chain shifts. In *Companion to phonology*, ed. Marc van Oostendorp, Beth Ewen, Colin Hume, and Keren Rice, 1717–1735. Wiley-Blackwell.
- Lubowicz, Anna. 2012. *The phonology of contrast*. Equinox publishing.
- Macken, Marlys. 1980. The child’s lexical representation: the ‘puzzle-puddle-pickle’ evidence. *Journal of Linguistics* 16.1:1–17.
- Magri, Giorgio. 2013. The complexity of learning in OT and its implications for the acquisition of phonotactics. *Linguistic Inquiry* 44.3:433–468.
- Magri, Giorgio. 2015. The learnability of  $\mathcal{F}$ -irrelevant phonotactic patterns. Manuscript.
- Magri, Giorgio. 2016a. Idempotency and chain shifts. In *Proceedings of WCCFL 33: the 33rd annual West Coast Conference in Formal Linguistics*, ed. Kyeong-min Kim. Cascadilla Proceedings Project.
- Magri, Giorgio. 2016b. On the definition of similarity orders in tesar’s (2013) theory of output-drivenness. Manuscript.
- Magri, Giorgio. to appeara. Idempotency in Optimality Theory. *Journal of Linguistics* .
- Magri, Giorgio. to appearb. Idempotency, output-drivenness and the faithfulness triangle inequality: some consequences of McCarthy’s (2013) categoricity generalization. *Journal of Logic, Language, and Information* .
- Magri, Giorgio. to appearc. Output-drivenness and partial phonological features. *Linguistic Inquiry* .
- Manzini, M. Rita, and Ken Wexler. 1987. Parameters, binding theory, and learnability. *Linguistic Inquiry* 18.3:413–444.
- McCarthy, John J. 1993. The parallel advantage: Containment, Consistency, and Alignment. Handout of talk presented at ROW-1, Rutgers University, New Brunswick, N.J.
- McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16:331–399.
- McCarthy, John J. 2002. Comparative markedness (long version). Ms. University of Amherst.
- McCarthy, John J. 2003a. Comparative markedness. *Theoretical linguistics* 29:1–51.
- McCarthy, John J. 2003b. OT constraints are categorical. *Phonology* 20:75–138.
- McCarthy, John J. 2004. *Optimality theory in phonology: A reader*. Oxford and Malden, MA: Blackwell.
- McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
- McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25:271–319.
- McCarthy, John J., and Alan Prince. 1995. Faithfulness and reduplicative identity. In *University of Massachusetts occasional papers in linguistics 18: Papers in optimality theory*, ed. Jill Beckman, Suzanne Urbanczyk, and Laura Walsh Dickey, 249–384. Amherst: GLSA.
- McLeod, S., J. van Doorn, and V. Reed. 2001. Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology* 10:99–110.
- Merchant, Nazarré. 2008. Discovering underlying forms: contrast pairs and ranking. Doctoral Dissertation, Rutgers University.
- Meyer, Marie-Christine. 2013. Ignorance and grammar. Doctoral Dissertation, MIT.
- Milsark, Gary L. 1977. Toward an explanation of certain peculiarities of the existential. *Linguistic Analysis* 3:1–29.
- Moreton, Elliott. 2004a. A compendium of synchronic chain shifts. Ms., Chapel Hill, NC: University of North Carolina Chapel Hill.
- Moreton, Elliott. 2004b. Non-computable functions in Optimality Theory. In *Optimality theory in phonology: A reader*, ed. John J. McCarthy, 141–163. Malden: MA: Wiley-Blackwell.
- Moreton, Elliott, and Paul Smolensky. 2002. Typological consequences of local constraint conjunction. In *WCCFL 21: Proceedings of the 21st annual conference of the West Coast Conference on Formal Linguistics*, ed. L. Mikkelsen and C Potts, 306–319. Cambridge, MA: Cascadilla Press.
- Orgun, C. Orhan. 1995. Correspondence and identity constraints in two-level Oprimality Theory. Ms., University of California, Berkeley.
- Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. In *The prosody-morphology interface*, ed. René Kager, Harry van der Hulst, and Wim Zonneveld, 310–343. Cambridge University Press. Reprinted in McCarthy (2004).
- Pater, Joe. 2008. Gradual learning and convergence. *Linguistic Inquiry* 39:334–345.
- Pater, Joe. 2009. Weighted constraints in Generative Linguistics. *Cognitive Science* 33:999–1035.
- Pater, Joe, and Jessica A. Barlow. 2003. Constraint conflict in cluster reduction. *Journal of Child Language* 30:487–526.



- Percus, Orin. 2006. Antipresuppositions. In *Theoretical and empirical studies of reference and anaphora: toward the establishment of generative grammar as an empirical science*, ed. Ayumi Ueyama. Report of the Grant-in-Aid for Scientific Research (B), Project No. 15320052, Japan Society for the Promotion of Science, 52-73.
- Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, and Michael Becker. 2010. Harmonic Grammar with Linear Programming: From linear systems to linguistic typology". *Phonology* 27(1):1-41.
- Prince, Alan. 2007. The pursuit of theory. In *The Cambridge handbook of phonology*, ed. Paul de Lacy, 33-60. Cambridge University Press.
- Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell. As Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993. Also available as ROA 537 version.
- Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. In *Constraints in phonological acquisition*, ed. R. Kager, J. Pater, and W. Zonneveld, 245-291. Cambridge University Press.
- Riggle, Jason. 2009. The complexity of ranking hypotheses in Optimality Theory. *Computational Linguistics* 35(1):47-59.
- Romoli, Jacopo. 2012. Obligatory scalar implicatures and relevance. *Snippets* 25:11-12. Available at <http://www.ledonline.it/snippets/allegati/snippets25004.pdf>.
- Rudin, Walter. 1953. *Principles of mathematical analysis*. McGraw-Hill Book Company.
- Sapir, Edward, and Morris Swadesh. 1978. Nootka texts.
- Sauerland, Uli. 2008. Implicated presuppositions. In *Sentence and context. language, context and cognition*, ed. Anita Steube. Berlin, Germany: Mouton de Gruyter.
- Schlenker, Philippe. 2012. "Maximize Presupposition" and Gricean reasoning. *Natural Language Semantics* 20:391-429.
- Schwarzschild, Roger. 1996. *Pluralities*. Dordrecht, Boston, London: Kluwer Academic Publisher.
- Schölkopf, Bernhard, and Alexander Smola. 2002. *Learning with kernels*. Cambridge, MA: MIT Press.
- Selkirk, Elisabeth. 1977. Some remarks on noun phrase structure. In *Formal syntax*, ed. Peter Culicover, Thomas Wasow, and Adrian Akmajian, 285-325. London: Academic Press.
- Shepard, Roger N, Carl I Hovland, and Herbert M Jenkins. 1961. Learning and memorization of classifications. *Psychological Monographs: General and Applied* 75:1-42.
- Singh, Raj. 2009. Maximize Presupposition! and informationally encapsulated implicatures. In *Proceedings of Sinn und Bedeutung 13*, ed. Arndt Riester and Torgrim Solstad, 513-526. Stuttgart.
- Singh, Raj. 2010. Oddness and ignorance inferences. Talk delivered at McGill University.
- Singh, Raj. 2012. Implicature cancellation and exhaustivity. In *Proceedings of SALT 22*.
- Smit, A., L. Hand, J. Freilinger, and A. Bernthal, J. Bird. 1990. The iowa articulation norms project and its nebraska replication. *Journal of Speech and Hearing Disorders* 55:779-798.
- Smith, N. V. 1973. *The acquisition of phonology: a case study*. Cambridge, England: CUP.
- Smolensky, Paul. 1995. On the internal structure of the constraint component of UG. URL <http://roa.rutgers.edu/article/view/87>, colloquium presented at the Univ. of California, Los Angeles, April 7, 1995. Handout available as ROA-86 from the Rutgers Optimality Archive.
- Smolensky, Paul, and G eraldine Legendre. 2006. *The harmonic mind*. Cambridge, MA: MIT Press.
- Spector, Benjamin. 2014. Scalar implicatures, blindness and common knowledge: Comments on magri (2011). In *Pragmatics, semantics and the case of scalar implicatures*, ed. Salvatore Pistoia Reda, 146-169. Palgrave Macmillan.
- Staroverov, Petr. 2014. Splitting theory and consonants epenthesis. Doctoral Dissertation, Rutgers University.
- Tesar, Bruce. 1995. Computational Optimality Theory. Doctoral Dissertation, University of Colorado, Boulder. Rutgers Optimality Archive ROA 90. Available at <http://www.roa.rutgers.edu>.
- Tesar, Bruce. 2013. *Output-driven phonology: Theory and learning*. Cambridge Studies in Linguistics.
- Tesar, Bruce, and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29:229-268.
- Thomas, Guillaume. 2012. Temporal implicatures. Doctoral Dissertation, MIT.
- Velten, Harry de Veltheyme. 1943. The growth of phonemic and lexical patterns in infant language. *Language* 19:281-292.

- Walker, Rachel. 1999. Esimbi vowel height shift: implications for faith and markedness. URL <http://roa.rutgers.edu/article/view/346>, ms, University of Southern California. Available as ROA-336 from the Rutgers Optimality Archive.
- Wheeler, Max W. 2005. Cluster reduction: Deletion or coalescence? *Catalan Journal of Linguistics* 4:57–82.
- White, James. 2013. Bias in phonological learning: Evidence from saltation. Doctoral Dissertation, UCLA.
- White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130.1:96–115.
- Wilson, Colin. 2001. Consonant cluster neutralization and targeted constraints. *Phonology* 18:147–197.
- Wolf, Matthew. 2007. What constraint connectives should be permitted in OT? *University of Massachusetts Occasional Papers in Linguistics* 36:151–179.
- Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: corpus and survey evidence from tagalog infixation. *Language* 83:277–316.
- Zuraw, Kie. 2013. \*MAP constraints. URL [www.linguistics.ucla.edu/people/zuraw/dnldpprs/star\\_map.pdf](http://www.linguistics.ucla.edu/people/zuraw/dnldpprs/star_map.pdf), unpublished manuscript, UCLA.



## APPENDIX A

### List of publications (as of June 2016)

#### A.1. Papers in journals

- [sub 4] ‘Output-drivenness and partial phonological features.’ Under revision for *Linguistic Inquiry*.
- [sub 3] ‘Idempotency, output-drivenness and the faithfulness triangle inequality: some consequences of McCarthy’s (2003) categoricity generalization.’ Submitted to the *Journal of Logic, Language, and Information*.
- [sub 2] ‘Idempotency in Optimality Theory.’ Submitted to the *Journal of Linguistics*.
- [pub 40] ‘Error-driven learning in OT and HG: a comparison.’ To appear in *Phonology*.
- [pub 39] 2015. ‘How to keep the HG weights non-negative: the truncated Perceptron reweighing rule.’ *Journal of Language Modeling*, 3.2: 345–375.
- [pub 38] ‘Noise robustness and stochastic tolerance of OT error-driven ranking algorithms.’ To appear in the *Journal of Logic and Computation*.
- [pub 37] 2016. ‘Two puzzles raised by oddness in conjunction.’ *Journal of Semantics*, 33.1: 1-17.
- [pub 36] 2014. ‘Tools for the robust analysis of error-driven ranking algorithms and their implications for modeling the child acquisition of phonotactics.’ *Journal of Logic and Computation* 24.1: 135-186
- [pub 35] 2013. ‘An argument for nominal Lexical Cumulativity.’ *Snippets* 27: 10-11.
- [pub 34] 2013. ‘HG has no computational advantages over OT: towards a new toolkit for computational OT.’ *Linguistic Inquiry* 44.4: 569–609.
- [pub 33] 2013. ‘The complexity of learning in Optimality Theory and its implications for the acquisition of phonotactics.’ *Linguistic Inquiry* 44.3: 434-468.
- [pub 32] 2013. ‘A note on the GLA’s choice of the current loser from the perspective of factorizability.’ *Journal of Logic, Language and Information* 22.2: 231-247.
- [pub 31] 2012. ‘Convergence of error-driven ranking algorithms.’ *Phonology* 29.2: 213-269.
- [pub 30] 2011. ‘The plurality inference of object mass nouns.’ *Snippets* 24: 9-10.
- [pub 29] 2011. ‘Another argument for embedded scalar implicatures based on oddness in downward-entailing environments.’ *Semantics & Pragmatics* 4.6: 1-51.
- [pub 28] 2009. ‘A Theory of individual-level predicates based on blind mandatory scalar implicatures.’ *Natural Language Semantics* 17.3: 245-297.

#### A.2. Books, book chapters, and book reviews

- [sub 1] ‘Blindness or short-sightedness? A reply to Schlenker (2012).’ Submitted to *Implicatures and Presuppositions*; Filippo Domaneschi and Salvatore Pistoia Reda (eds.); Palgrave Macmillan.
- [pub 27] 2015. ‘Restrictiveness through relaxation.’ In *Short’ schrift for Alan Prince*; Eric Bakovic (ed.).
- [pub 26] 2015. ‘Review of Tesar’s (2013) *Output-driven phonology: Theory and learning*.’ *Phonology* 31: 525-556.
- [pub 25] 2014. ‘An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening.’ In *Pragmatics, Semantics and the Case of Scalar Implicatures*; Salvatore Pistoia Reda (ed.); Palgrave Studies in Pragmatics, Language and Cognition; Palgrave Macmillan, Basingstoke; pp. 99–145.
- [pub 24] 2009. ‘A Theory of Individual-Level Predicates Based on Blind Mandatory Scalar Implicatures. Constraint Promotion for Optimality Theory.’ Published by MITWPL (MIT Working Papers in Linguistics).

### A.3. Papers in proceedings of peer reviewed conferences

- [pub 23] 2015. ‘How to choose successful losers in error-driven phonotactic learning.’ Joint work with René Kager. In **MoL 14: Proceedings of the 14th meeting on the Mathematics of Language**; eds. Marco Kuhlmann, Makoto Kanazawa, and Gregory M. Kobele; pp. 126-138; Association for Computational Linguistics.
- [pub 22] 2015. ‘Universals on natural language determiners from a PAC-learnability perspective.’ In **CogSci 37: Proceedings of the 37th Annual Conference of the Cognitive Science Society**; eds. D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, and P. P. Maglio; pp. 1494-1499; Austin, TX: Cognitive Science Society.
- [pub 21] 2015. ‘Idempotency and chain shifts.’ In **WCCFL 33: Proceedings of the 33rd West Coast Conference on Formal Linguistics**; eds. Kyeong-min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson, and Lisa Shorten; Cascadilla Proceedings Project.
- [pub 20] 2015. ‘How to account for phonetically counterintuitive segment inventories using only phonetically grounded markedness constraints.’ Joint work with René Kager. In **NELS 45: Proceedings of the 45th annual meeting of the North East Linguistic Society**; ed. Thuy Bui and Deniz Özyildiz; pp. 145-154; GLSA Publications.
- [pub 19] 2014. ‘The error-driven ranking model of the acquisition of phonotactics: how to keep the faithfulness constraints at bay.’ In **MORPHFSM 2014: Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM**; eds. Özlem Çetinoglu, Jeffrey Heinz, Andreas Maletti, and Jason Riggle; Association for Computational Linguistics.
- [pub 18] 2014. ‘Error-driven versus batch models of the acquisition of phonotactics: David defeats Goliath.’ In **AMP1: Supplemental Proceedings of the Annual Meeting in Phonology 2013**; eds. John Kingston, Claire Moore-Cantwell, Joe Pater, and Robert Staubs; Linguistic Society of America.
- [pub 17] 2013. ‘A closer look at Boersma and Hayes’ (2001) Ilokano metathesis test case.’ Joint work with Benjamin Storme. To appear in **CLS 49: Proceedings of the 49th annual conference of the Chicago Linguistics Society**.

- [pub 16] 2013. ‘The error-driven ranking model of the early stage of the acquisition of phonotactics: an initial result on restrictiveness.’ In **NELS 43: Proceedings of the 43rd annual meeting of the North East Linguistic Society**; eds. Hsin-Lun Huang, Ethan Poole, and Amanda Rysling; GLSA Publications. pp. 277-290.
- [pub 15] 2013. ‘Constraint promotion: not only convergent but also efficient.’ In **CLS 48: Proceedings of the 48th annual conference of the Chicago Linguistics Society**.
- [pub 14] 2013. ‘No need for a theory of the distribution of readings of English bare plurals.’ In **SALT 22: Proceedings of the 22nd annual conference Semantics and Linguistic Theory**; ed. Anca Chereches; pp. 383-402; Linguistic Society of America.
- [pub 13] 2012. ‘An approximation approach to the problem of the acquisition of phonotactics in Optimality Theory.’ In **ACL-SIGMORPHON 12: Proceedings of the 12th meeting of the ACL Special Interest Group in Morphology and Phonology**; eds. Lynne Cahill and Adam Albright; pp. 52-61; Association for Computational Linguistics.
- [pub 12] 2012. ‘Collective nouns without groups.’ In **IATL 27: Proceedings of the 27th annual meeting of the Israeli Association for Theoretical Linguistics**; ed. Evan Cohen; pp. 183-202; MIT Working Papers in Linguistics #65.
- [pub 11] 2011. ‘HG has no computational advantages over OT.’ In **WCCFL 29: Proceedings of the 29th West Coast Conference on Formal Linguistics**; eds. J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, and A. Trueman; pp. 380-388; Cascadilla Proceedings Project.
- [pub 10] 2011. ‘An online model of the acquisition of phonotactics within Optimality Theory.’ In **CogSci 33: Expanding the space of cognitive science: Proceedings of the 33rd annual meeting of the Cognitive Science Society**; eds. L. Carlson, C. Hölscher, and T. Shipley; pp. 2012-2017; Cognitive Science Society.
- [pub 9] 2011. ‘Correctness of promotion/demotion OT online models on some case studies by Prince & Tesar 2004.’ In **CLS 47: Proceedings of the 47th annual conference of the Chicago Linguistics Society**.
- [pub 8] 2011. ‘Another argument for embedded scalar implicatures based on oddness in downward entailing contexts.’ In **SALT 20: Proceedings of the 20th annual conference Semantics And Linguistic Theory**; eds. N. Li and D. Lutz; pp. 564-581.
- [pub 7] 2010. ‘On the complexity of the problem of the acquisition of phonotactics in Optimality Theory.’ In **ACL-SIGMORPHON 11: Proceedings of the 11th meeting of the ACL Special Interest Group in Morphology and Phonology**; eds. J. Heinz, L. Cahill, and R. Wicentowski; pp. 19-27; Association for Computational Linguistics
- [pub 6] 2009. ‘Mismatching scalar implicatures.’ In *Presuppositions and Implicatures: Proceedings of the MIT-Paris Workshop*; ed. P. Egré and G. Magri; pp. 153-168; MIT Working Papers in Linguistics #60.
- [pub 5] 2007. ‘Modeling doubly marked lags with a split additive model.’ Joint work with A. Albright and J. Michaels. In **BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development**; eds. H. Chan, H. Jacob, and E. Kapla; vol. I, pp. 36-47; Cascadilla Press.
- [pub 4] 2007. ‘Quadratic phonotactics.’ In **NELS 38: Proceedings of the 38th annual meeting of the North East Linguistic Society**; eds. A. Schardl, M. Walkow, and M. Abdurrahman; GLSA Publications.
- [pub 3] 2007. ‘The sortal theory of plurals.’ In **SuB 12: Proceedings of the 12th annual conference Sinn und Bedeutung**; ed. A. Grønn; pp. 399-413.

- [pub 2] 2006. ‘The blindness hypothesis and individual-level predicates.’ In **SALT 16: Proceedings of the 16th annual conference Semantics and Linguistic Theory**; eds. M. Gibson and J. Howell; pp. 102-119; CLC Publications.
- [pub 1] 2004. ‘Constraints on the existential reading of bare plural subjects of ILPs: syntax or semantics?’ In **NELS 35: Proceedings of the 35th annual meeting of the North East Linguistic Society**; eds. L. Bateman and C. Ussery; vol. I, pp. 391-402; GLSA Publications.

#### A.4. Manuscripts (available at <https://sites.google.com/site/magrigr/>)

- [man 4] ‘Mass nouns, *only* and the logicity of the minimum operator in Natural Language Semantics.’ To be submitted to **Synthese**.
- [man 3] ‘On the definition of similarity orders in the theory of output-drivenness. A note on Tesar (2013).’
- [man 2] ‘A MaxSat formulation of the Subset problem in Optimality Theory.’ To be submitted to **Transactions of the Association for Computational Linguistics**.
- [man 1] ‘When is it the case that one tableau suffices? A note on Prince (2015)?’ Also available on the Rutgers Optimality Archive as manuscript ROA-1253. To be submitted to the **Journal of Language Modeling**.

#### A.5. Manuscripts in preparation

- [in prep 4] ‘Child vowel epenthesis in the INANP database.’ Joint work with Adam Albright. This paper is based on results presented at ICPC (International Child Phonology Conference) and at GLOW (Generative Grammar in the Old World Conference) in 2013.
- [in prep 3] ‘Restrictiveness and symmetric faithfulness constraints.’ To be submitted to a special issue of **Phonology** on computational phonology in March 2016.
- [in prep 2] ‘Some restrictiveness results for the calibrated error-driven ranking algorithm.’ To be submitted to **Natural Language and Linguistic Theory**. This paper is based on results presented in [pub 16], [pub 18], [pub 20], and [pub 23].
- [in prep 1] ‘How to keep the faithfulness constraints at bay.’ To be submitted to **Computational Linguistics**. This paper is based on results presented in [pub 19].

## APPENDIX B

### Curriculum vitae (as of June 2016)

#### B.1. Contact information

UMR 7023 SFL (CNRS and Université Paris VIII)      0033 (0)7 81 36 41 95  
59/61 rue Pouchet      magrigrg@gmail.com  
75017 Paris, France  
<https://sites.google.com/site/magrigrg/>

#### B.2. Employment

2012–present    CNRS researcher (CR1)  
                  UMR 7023 SFL (CNRS and University of Paris VIII), Paris  
2013-2015      Visiting researcher (on a Marie Curie Intra-European Fellowship  
                  for Career Development),  
                  Utrecht Institute of Linguistics OTS (UiL OTS), Utrecht  
2012            Post-doc (Labex Empirical Foundations of Linguistics)  
                  UMR 7110 LLF (CNRS and University of Paris VII), Paris  
2009 – 2012    Post-doc  
                  Cognitive Science Department (DEC), École Normale Supérieure, Paris

#### B.3. Education

2011            18th Machine Learning Summer School  
                  Bordeaux, September 4-17  
2004 – 2009    Ph.D in Linguistics  
                  (five year program; double major: Semantics and Computational Phonology)  
                  Department of Linguistics and Philosophy, MIT, USA  
2003 – 2004    Visiting student  
                  Department of Linguistics and Philosophy, MIT, USA  
2002 – 2004    First two years of the Cognitive Neuroscience Ph.D Program  
                  Department of Psychology, University of Milano Bicocca, Italy  
2001 – 2002    “Laurea” in Mathematics (four year program; summa cum laude)  
                  Department of Mathematics, University of Milano, Italy  
1994 – 2001    “Laurea” in Philosophy (four year program; summa cum laude)  
                  Department of Philosophy, University of Milano, Italy

#### B.4. PhD dissertation

*institution:*    Department of Linguistics and Philosophy, MIT  
*date of award:*    September 2009  
*title:*        *A Theory of individual-level predicates based on blind mandatory scalar implicatures. Constraint promotion for Optimality Theory*  
*advisors:*      Adam Albright and Danny Fox (co-chairs)  
*committee:*     Mark Johnson (external member: Brown University); Irene Heim;  
                  Donca Steriade

#### B.5. Grants



- 2015 *MIT France Seed Fund Award*. Project title: ‘Phonological Typology and Learnability’ (19,000\$). With Donna Steriade (co-PI). Hosted at the Department of Linguistics and Philosophy, MIT
- 2011 *Fyssen Research Grant*. Project title: ‘Child acquisition of sound patterns: a computational modeling approach’ (9.500€). Hosted at the UMR 7023 SFL (CNRS, University of Paris VIII).
- 2011 *Marie Curie Intra-European Fellowship for Career Development* (FP7-PEOPLE-2011-IEF). Project title: ‘Modeling the acquisition of phonotactics’ (191,675€). Hosted at UiL-OTS, Utrecht University.
- 2004 MIT Presidential Fellowship (awarded to “the most outstanding students worldwide to pursue graduate studies at the Institute”).

### B.6. Papers in journals

- ‘Idempotency, output-drivenness and the faithfulness triangle inequality: some consequences of McCarthy’s (2003) categoricity generalization.’ Submitted to the **Journal of Logic, Language, and Information**.
- ‘Idempotency in Optimality Theory.’ Submitted to the **Journal of Linguistics**.
- ‘Output-drivenness and partial phonological features.’ Under revision for **Linguistic Inquiry**.
- to appear ‘Error-driven learning in OT and HG: a comparison.’ To appear in **Phonology**.
- to appear ‘Noise robustness and stochastic tolerance of OT error-driven ranking algorithms.’ To appear in the **Journal of Logic and Computation**.
- 2016 ‘Two puzzles raised by oddness in conjunction.’ **Journal of Semantics**, 33.1: 1-17.
- 2015 ‘How to keep the HG weights non-negative: the truncated Perceptron reweighing rule.’ **Journal of Language Modeling**, 3.2: 345–375.
- 2014 ‘Tools for the robust analysis of error-driven ranking algorithms and their implications for the acquisition of phonotactics.’ **Journal of Logic and Computation** 24.1: 135-186
- 2013 ‘An argument for nominal Lexical Cumulativity.’ **Snippets** 27: 10-11.
- 2013 ‘HG has no computational advantages over OT: towards a new toolkit for computational OT.’ **Linguistic Inquiry** 44.4: 569–609.
- 2013 ‘The complexity of learning in Optimality Theory and its implications for modeling the child early acquisition of phonotactics.’ **Linguistic Inquiry** 44.3: 434-468.
- 2013 ‘A note on the GLA’s choice of the current loser from the perspective of factorizability.’ **Journal of Logic, Language and Information** 22.2: 231-247.
- 2012 ‘Convergence of error-driven ranking algorithms.’ **Phonology** 29.2: 213-269.
- 2011 ‘The plurality inference of object mass nouns.’ **Snippets** 24: 9-10.
- 2011 ‘Another argument for embedded scalar implicatures based on oddness in downward-entailing environments.’ **Semantics and Pragmatics** 4.6: 1-51.
- 2009 ‘A Theory of individual-level predicates based on blind mandatory scalar implicatures.’ **Natural Language Semantics** 17.3: 245-297.

### B.7. Books, book chapters, and book reviews

- ‘Blindness or Short-sightedness? A reply to Schlenker (2012).’ Submitted to *Implicatures and Presuppositions*; Filippo Domaneschi and Salvatore Pistoia Reda (eds.); Palgrave Macmillan.
- 2015 ‘Restrictiveness through relaxation.’ In *Short’s schrift for Alan Prince*; Eric Bakovic (ed.).
- 2015 ‘Review of Tesar’s (2013) *Output-driven phonology: Theory and learning*.’ *Phonology* 31: 525-556.
- 2014 ‘An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening.’ In *Pragmatics, Semantics and the Case of Scalar Implicatures*; Salvatore Pistoia Reda (ed.); Palgrave Studies in Pragmatics, Language and Cognition; Palgrave Macmillan, Basingstoke; pp. 99–145.
- 2009. ‘A Theory of Individual-Level Predicates Based on Blind Mandatory Scalar Implicatures. Constraint Promotion for Optimality Theory.’ Published by MITWPL.

#### B.8. Manuscripts (available at <https://sites.google.com/site/magrigrg/>)

- 2015 ‘Mass nouns, *only*, and the logicity of the minimum operator in Natural Language Semantics.’ To be submitted to **Synthese**.
- 2015 ‘On the definition of similarity orders in the theory of output-drivenness. A note on Tesar (2013).’
- 2015 ‘A MaxSat formulation of the Subset problem in Optimality Theory.’ To be submitted to the **Transactions of the Association for Computational Linguistics**.
- 2015 ‘When is it the case that one tableau suffices? A note on Prince (2015)?’ Also available on the Rutgers Optimality Archive as ROA-1253. To be submitted to the **Journal of Language Modeling**.

#### B.9. Papers in proceedings of peer reviewed conferences

- 2016 ‘Idempotency and chain shifts.’ In *WCCFL 33: Proceedings of the 33rd West Coast Conference in Formal Linguistics*; eds. Kyeong-min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson, and Lisa Shorten; Cascadilla Proceedings Project.
- 2015 ‘How to choose successful losers in error-driven phonotactic learning.’ Joint work with René Kager. In *MoL 14: Proceedings of the 14th meeting on the Mathematics of Language*; eds. Marco Kuhlmann, Makoto Kanazawa, and Gregory M. Kobele; pp. 126-138; Association for Computational Linguistics.
- 2015 ‘Universals on natural language determiners from a PAC-learnability perspective.’ In *CogSci 37: Proceedings of the 37th Annual Conference of the Cognitive Science Society*; eds. D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, and P. P. Maglio; pp. 1494-1499; Austin, TX: Cognitive Science Society.
- 2015 ‘How to account for phonetically counterintuitive segment inventories using only phonetically grounded markedness constraints.’ Joint work with René Kager. In *NELS 45: Proceedings of the 45th annual meeting of the North East Linguistic Society*; ed. Thuy Bui and Deniz Özyildiz; pp. 145-154; GLSA Publications.
- 2014 ‘The error-driven ranking model of the acquisition of phonotactics: how to keep the faithfulness constraints at bay.’ In *MORPHFSM 2014: Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM*; eds. Özlem Çetinoglu, Jeffrey Heinz, Andreas Maletti, and Jason Riggle; Association for Computational Linguistics.
- 2014 ‘Error-driven versus batch models of the acquisition of phonotactics: David defeats Goliath.’ In *Supplemental Proceedings of Phonology 2013*; eds. John Kingston, Claire Moore-Cantwell, Joe Pater, and Robert Staubs; Linguistic Society of America.

- 2013 ‘A closer look at Boersma and Hayes’ (2001) Ilokano metathesis test case.’ Joint work with Benjamin Storme. To appear in *CLS 49: Proceedings of the 49th annual conference of the Chicago Linguistics Society*.
- 2013 ‘The error-driven ranking model of the early stage of the acquisition of phonotactics: an initial result on restrictiveness.’ In *NELS 43: Proceedings of the 43rd annual meeting of the North East Linguistic Society*; eds. Hsin-Lun Huang, Ethan Poole, and Amanda Rysling; GLSA Publications. pp. 277-290.
- 2013 ‘Constraint promotion: not only convergent but also efficient.’ In *CLS 48: Proceedings of the 48th annual conference of the Chicago Linguistics Society*.
- 2013 ‘No need for a theory of the distribution of readings of English bare plurals.’ In *SALT 22: Proceedings of the 22nd annual conference Semantics and Linguistic Theory*; ed. Anca Chereches; pp. 383-402; Linguistic Society of America.
- 2012 ‘An approximation approach to the problem of the acquisition of phonotactics in Optimality Theory.’ In *ACL-SIGMORPHON 12: Proceedings of the 12th meeting of the ACL Special Interest Group in Morphology and Phonology*; eds. Lynne Cahill and Adam Albright; pp. 52-61; Association for Computational Linguistics.
- 2012 ‘Collective nouns without groups.’ In *IATL 27: Proceedings of the 27th annual meeting of the Israeli Association for Theoretical Linguistics*; ed. Evan Cohen; pp. 183-202; MIT Working Papers in Linguistics #65.
- 2011 ‘HG has no computational advantages over OT.’ In *WCCFL 29: Proceedings of the 29th West Coast Conference on Formal Linguistics*; eds. J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, and A. Trueman; pp. 380-388; Cascadilla Proceedings Project.
- 2011 ‘An online model of the acquisition of phonotactics within Optimality Theory.’ In *CogSci 33: Expanding the space of cognitive science: Proceedings of the 33rd annual meeting of the Cognitive Science Society*; eds. L. Carlson, C. Hölscher, and T. Shipley; pp. 2012-2017; Cognitive Science Society.
- 2011 ‘Correctness of promotion/demotion OT online models on some case studies by Prince & Tesar 2004.’ In *CLS 47: Proceedings of the 47th annual conference of the Chicago Linguistics Society*.
- 2011 ‘Another argument for embedded scalar implicatures based on oddness in downward entailing contexts.’ In *SALT 20: Proceedings of the 20th annual conference Semantics And Linguistic Theory*; eds. N. Li and D. Lutz; pp. 564-581.
- 2010 ‘On the complexity of the problem of the acquisition of phonotactics in Optimality Theory.’ In *ACL-SIGMORPHON 11: Proceedings of the 11th meeting of the ACL Special Interest Group in Morphology and Phonology*; eds. J. Heinz, L. Cahill, and R. Wicentowski; pp. 19-27; Association for Computational Linguistics
- 2009 ‘Mismatching scalar implicatures.’ In *Presuppositions and Implicatures: Proceedings of the MIT-Paris Workshop*; ed. P. Egré and G. Magri; pp. 153-168; MIT Working Papers in Linguistics #60.
- 2007 ‘Modeling doubly marked lags with a split additive model.’ Joint work with A. Albright and J. Michaels. In *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development*; eds. H. Chan, H. Jacob, and E. Kapla; vol. I, pp. 36-47; Cascadilla Press.
- 2007 ‘Quadratic phonotactics.’ In *NELS 38: Proceedings of the 38th annual meeting of the North East Linguistic Society*; eds. A. Schardl, M. Walkow, and M. Abdurrahman; GLSA Publications.
- 2007 ‘The sortal theory of plurals.’ In *SuB 12: Proceedings of the 12th annual conference Sinn und Bedeutung*; ed. A. Grønn; pp. 399-413.

- 2006 ‘The blindness hypothesis and individual-level predicates.’ In *SALT 16: Proceedings of the 16th annual conference Semantics and Linguistic Theory*; eds. M. Gibson and J. Howell; pp. 102-119; CLC Publications.
- 2004 ‘Constraints on the existential reading of bare plural subjects of ILPs: syntax or semantics?’ In *NELS 35: Proceedings of the 35th annual meeting of the North East Linguistic Society*; eds. L. Bateman and C. Ussery; vol. I, pp. 391-402; GLSA Publications.

### B.10. Talks and posters in peer reviewed conferences

- [2016] ‘Idempotence and the triangle inequality: some consequences of McCarthy’s categoricity generalization.’ Talk to be presented at the 13th annual Old World Conference in Phonology (**OCP 13**), Budapest, 13-16 January.
- 2015 ‘How to choose successful losers in error-driven phonotactic learning.’ Joint work with René Kager. Presented at the 14th meeting on the Mathematics of Language (**MoL 14**), Chicago, 25-26 July.
- 2015 ‘Universals on natural language determiners from a PAC-learnability perspective.’ Presented at the 37th annual conference of the Cognitive Science Society (**CogSci 37**), Pasadena, 23-25 July.
- 2015 ‘Idempotency in OT.’ Presented at the 33rd West Coast Conference on Formal Linguistics (**WCCFL 33**), Vancouver, 28 March.
- 2015 ‘Idempotency and the problem of the acquisition of phonotactics.’ Presented at the 12th Old World Phonology conference (**OCP 12**), Barcelona, 29 January.
- 2014 ‘The error-driven ranking model learns inventories of obstruents.’ Joint work with Rene Kager. Poster presented at the 45th annual meeting of the North East Linguistic Society (**NELS 45**), MIT, 1 November.
- 2014 ‘The error-driven ranking model and the problem of  $\mathcal{F}$ -control.’ Presented at the Joint Meeting of SIGMORPHON and SIGFSM (**MORPHFSM 2014**), Baltimore, 27 June.
- 2014 ‘On the Prince-Tesar-Hayes’ approach to restrictiveness in Optimality Theory.’ Presented at the 22nd Manchester Phonology Meeting (**MFM 22**), Manchester, 29-31 May.
- 2014 ‘Error-driven learning in HG: not impossible, but definitely not optimal.’ Presented at the 11th Old World Conference in Phonology (**OCP 11**), Leiden and Amsterdam, 25 January.
- 2013 ‘Error-driven versus batch models of the acquisition of phonotactics: David defeats Goliath.’ Poster presented at the 1st Phonology Conference (**AMP 2013**), UMASS, Amherst, 10 November.
- 2013 ‘Error-driven versus batch models of the early stage of the acquisition of phonotactics: David defeats Goliath.’ Presented at the 11th French Phonology Meeting (**RFP 11**), University of Nantes, 3 July.
- 2013 ‘The error-driven ranking model of the child’s acquisition of phonotactics.’ Poster presented at the Workshop on Infant Language Development (**WILD**), Basque Center for Cognition, Brain, and Language (BCBL), San Sebastian, Spain, 20 June.
- 2013 ‘Epenthesis in child phonology.’ Joint work with Adam Albright. Poster presented at the International Child Phonology Conference (**ICPC 2013**), Nijmegen, 12 June.
- 2013 ‘Deterministic and stochastic error-driven learning: a comparison between OT and HG.’ Poster presented at the 21st Manchester Phonology Meeting (**MFM 21**), Manchester, 25 May.

- 2013 ‘A closer look at Boersma and Hayes’ (2001) simulations.’ Joint work with Benjamin Storme. Presented at the 49th annual conference of the Chicago Linguistics Society (**CLS 49**), Chicago University, 20 April.
- 2013 ‘Perceptually motivated epenthesis asymmetries in the acquisition of consonant clusters.’ Joint work with Adam Albright. Presented at the 36th annual conference in Generative Grammar in the Old World (**GLOW 36**), Centre for Language and Literature, Lund University, 5 April.
- 2012 ‘The CEDRA model of the acquisition of phonotactics: four results and a conjecture.’ Presented at the Computational Linguistics workshop of the 43rd annual meeting of the North East Linguistic Society (**NELS 43**), City University of New York (CUNY), October 21.
- 2012 ‘The early stage of the acquisition of phonotactics from a learnability perspective.’ Presented at the 10th Meeting of the French Phonology Network (**RFP 10**), Paris, June 26.
- 2012 ‘An approximation approach to the problem of the acquisition of phonotactics in Optimality Theory.’ Presented at the 12th meeting of the ACL Special Interest Group in Morphology and Phonology (**ACL-SIGMORPHON 12**), Montreal, June 9.
- 2012 ‘No need for a theory of the distribution of readings of English bare plurals.’ Poster presented at the 22nd annual conference Semantics and Linguistic Theory (**SALT 22**), Chicago, May 18-20.
- 2012 ‘Constraint promotion: not only convergent but also efficient.’ Presented at the 48th annual conference of the Chicago Linguistics Society (**CLS 48**), Chicago, April 19-21.
- 2012 ‘No need for a theory of the distribution of readings of English bare plurals.’ Poster presented at the 22nd Conference on Generative Grammar (**CGG 22**), Barcellona, March 21-23.
- 2012 ‘No need for a theory of the distribution of readings of English bare plurals.’ Presented at the workshop “Languages with and without articles” (**LSALAA**), University of Paris 8, March 15-16.
- 2012 ‘A result on restrictiveness of the error-driven ranking model.’ Presented at the workshop “Psyco-computational Models of Human Language Acquisition” (**PsychoCompLA-2012**), Portland, January 5.
- 2012 ‘On the optimality of Optimality Theory.’ Presented at the 86th Annual Meeting of the Linguistic Society of America (**LSA 86**), Portland, January 8.
- 2011 ‘Online learning of lexicographic orders.’ Poster presented at the 18th Machine Learning Summer School (**MLSS 18**), Bordeaux, September 15.
- 2011 ‘An online model of the acquisition of phonotactics within Optimality Theory.’ Poster presented at the 33th annual meeting of the Cognitive Science Society (**CogSci 33**), Boston, July 19-24.
- 2011 ‘HG has no computational advantages over OT.’ Presented at the 29th West Coast Conference on Formal Linguistics (**WCCFL 29**), University of Arizona, April 23.
- 2011 ‘Correctness of promotion/demotion OT online models on some case studies by Prince & Tesar 2004.’ Presented at the 47th annual conference of the Chicago Linguistics Society (**CLS 47**), University of Chicago, April 8.
- 2011 ‘Strong correctness of OT online models for the acquisition of inventory segmental phonotactics.’ Presented at 85th annual meeting of the Linguistic Society of America (**LSA 85**), Pittsburgh, January 7.

- 2010 ‘On the Complexity of the Problem of the Acquisition of Phonotactics within Optimality Theory.’ Presented at the 11th meeting of the ACL Special Interest Group in Morphology and Phonology (**ACL/SIGMORPHON 11**), University of Uppsala, July 15.
- 2010 ‘Another argument for embedded scalar implicatures.’ Poster presented at the 20th annual conference Semantics and Linguistic Theory (**SALT 20**), UBC/SFU, April 29 - May 1.
- 2010 ‘Another argument for embedded scalar implicatures.’ Presented at the 36th Incontro di Grammatica Generativa (**IGG 36**), University of Milano-Bicocca, February 26-27.
- 2010 ‘Constraint promotion and the OT online model of the “early stage” of the acquisition of phonotactics.’ Presented at the workshop “Computational Modeling of Sound Pattern Acquisition” (**CMSPA**), University of Alberta, February 12-13.
- 2010 ‘Constraint promotion and the OT online model of the “early stage” of the acquisition of phonotactics.’ Presented at 84th annual meeting of the Linguistic Society of America (**LSA 84**), Baltimore, January 9.
- 2009 ‘A convergent version of the GLA that performs both promotion and demotion.’ Presented at the 31th annual meeting of the Deutsche Gesellschaft für Sprachwissenschaft (**DGfS 31**), Osnabrück, March 4-6.
- 2009 ‘On the acquisition of Dutch syllable types.’ Poster presented at 83rd annual meeting of the Linguistic Society of America (**LSA 83**), San Francisco, January 10.
- 2007 ‘Modeling doubly marked lags with a split additive model.’ Joint work with Adam Albright and Jennifer Michaels. Presented at the 32nd annual Boston University Conference on Language Development (**BUCLD 32**), Boston University, November 2-4.
- 2006 ‘The Blindness Hypothesis and individual-level predicates.’ Presented at the 16th annual conference Semantics and Linguistic Theory (**SALT 16**), University of Tokyo, March 22-24.

### B.11. Invited talks at conferences and workshops

- 2015 ‘Idempotency, output-drivenness and the faithfulness triangle inequality: some consequences of McCarthy’s (2003) categoricity conjecture.’ Invited talk at the *LSA Workshop on Computational Phonology and Morphology*, University of Chicago, July 11.
- 2015 ‘Comments on unarity.’ Invited commentary at the *Workshop on the Formal Structure of OT Typologies*, Rutgers University, May 29-30.
- 2013 ‘Computational modeling in Generative Grammar.’ Invited talk at the CNRS GDR 3195 ADYLOC (Acquisition, Dysfonctionnements, Langues, Oral, Cognition), Paris, 10 April.
- 2012 ‘Restrictiveness of error-driven ranking algorithms.’ Invited talk at the workshop *Phonotactic Grammar: Theories and Models*, Scuola Normale Superiore, Cortona, Italy, September 28-29.
- 2012 ‘The CEDRA model of the acquisition of phonotactics: three results and a conjecture.’ Invited presentation at *EALING* (Ecole d’Automne de Linguistique; Fall School in Linguistics), École Normale Supérieure, Paris, September 16.
- 2011 ‘A comparison between HG and OT from a computational perspective.’ Invited talk at the workshop *Optimality Theory as a General Cognitive Architecture* at *CogSci 33: the 33rd annual meeting of the Cognitive Science Society*, Boston, 20 July.
- 2011 ‘Commentary on S. Yalcin’s paper *Credal Expressivism*.’ Invited commentary at the *Semantics and Philosophy of Language Workshop*, All Souls College, Oxford, January 13-14 January.

- 2010 ‘An account for restrictions on bare plural subjects of ILPs based on blind mandatory scalar implicatures.’ Invited talk at the conference *Genericity: Interpretation and Uses II*, École Normale Supérieure, Paris; June 24-25.
- 2009 ‘An implicature-based theory of ILPs: the case of tense modifiers and tense morphology.’ Invited talk at the *Workshop on the ILP/SLP distinction* organized by Angelika Kratzer, Maison Suger, Paris, November 14.
- 2007 ‘Oddness is asymmetric.’ Invited talk at the *MIT-Paris Workshop on Presuppositions and Implicatures*, DEC-ENS, May 30-31.

### B.12. Other invited talks

- 2015 ‘Idempotency, output-drivenness, and the faithfulness triangle inequality.’ Rutgers Optimality Research Group (RORG), Rutgers University, Department of Linguistics, 18 September.
- 2014 ‘The error-driven ranking model of the acquisition of phonotactics: an overview.’ IJN, École Normale Supérieure, Paris, 29 January.
- 2012 ‘The OT error-driven ranking model of the acquisition of phonotactics.’ *Language Cognition and Computation seminar*, Chicago University, May 18.
- 2009 ‘An account for restrictions on bare plural subjects of ILPs based on blind mandatory scalar implicatures.’ Department of Linguistics, Oxford University (job talk), March 26.
- 2009 ‘A theory of individual-level predicates based on blind mandatory scalar implicatures.’ Department of Linguistics, Stanford University, department colloquium (job talk), January 30.
- 2009 ‘A new look at the Ranking Problem in Optimality Theory.’ Department of Linguistics, Stanford University, department seminar (job talk), January 29.
- 2007 ‘The theory of individual-level predicates and blind scalar implicatures.’ *Theory Group*, Harvard University, April 30.

### B.13. Contributed talks

- 2012 ‘Restrictiveness of OT error-driven ranking algorithms.’ *Phonology Circle*, MIT April 18.
- 2011 ‘The OT error-driven ranking model of the acquisition of phonotactics.’ *Ling-Lunch*, University of Paris 7, November 24.
- 2009 ‘A theory of individual-level predicates based on blind mandatory scalar implicatures.’ *Language and Cognition reading group*, Harvard University, October 13 and 20.
- 2009 ‘How to extend Tesar and Smolensky’s analysis of Constraint Demotion to Constraint Demotion/Promotion.’ *NECPhon 3: the 3rd North East Computational Phonology meeting*, MIT, October 24.
- 2008 ‘A Convergent version of the GLA for standard Optimality Theory.’ *NECPhon 2: the 2nd North East Computational Phonology meeting*, Yale University, November 15.
- 2007 ‘The multiplicative GLA.’ *NECPhon 1: the 1st North East Computational Phonology meeting*, University of Amherst, November 10.

### B.14. Teaching

- [2016] *An introduction to computational phonology/II: learning underlying forms*. Two week mini-course. Department of Linguistics and Philosophy, MIT.

- [2016] *An introduction to computational phonology/I: consistency and restrictiveness*. Two week mini-course. Department of Linguistics and Philosophy, MIT.
- [2016] *Computational constraint-based phonology: an introduction*. ACTL (Advanced Core Training in Linguistics) Summer School, UCL (<http://www.actl.ucl.ac.uk/summer.html>)
- 2015 *The error-driven ranking model of the acquisition of phonotactics*. Course at the LSA Linguistics Summer Institute, Chicago (<https://lsa2015.uchicago.edu/>)
- 2014 *Phonology: Typology and Acquisition*. Utrecht University, LIMV13010. Master-level course co-taught with Aoju Chen and René Kager.
- 2012 *Cognitive Computational Phonology: the case study of error-driven models of the acquisition of phonotactics*. 24th ESSLLI: European Summer School in Logic, Language and Information. Opole, Poland; 13-17 August (<http://esslli2012.pl/index.php?id=132>).
- 2011 *Formal methods for linguistics* (DEC-B06-S1). Master-level course. Instructor.
- 2011 *Semantics and Pragmatics: Formal and experimental approaches* (DEC-B17-S2). Master class taught by Emmanuel Chemla and Vincent Homer. TA (duties: instructing weekly sections).
- 2010 ‘The OT online model of early stage of the acquisition of phonotactics.’ Advanced mini-course (4 classes) taught at *EALing: Ecole d’Automne de Linguistique*, École Normale Supérieure.
- 2010 *Semantics and Pragmatics: Formal and experimental approaches* (DEC-B17-S2). Master class taught by Philippe Schlenker and Emmanuel Chemla. TA (duties: instructing weekly sections).
- 2007 *Advanced Semantics* (MIT 24.973). Graduate class co-taught with Irene Heim.
- 2006 *Introduction to Linguistics* (MIT 24.900). Undergraduate class taught by Donca Steriade. TA (duties: grading and instructing weekly sections).
- 2005 *Introduction to Semantics* (MIT 24.970). Graduate class taught by Kai von Fintel. TA (duties: instructing weekly sections).

### B.15. Conference organization

- [2016] Fringe meeting of the *MF24: the 24th Manchester Phonology Conference*; co-organizer with Jeffrey Heinz; University of Manchester, 25 May (<https://sites.google.com/site/24mfmfringe/home>)
- 2015 GLOW phonology workshop *The implications of computation and learnability for phonological theory*; co-organizer; Paris, April 18 (<https://sites.google.com/site/2015glow/home/workshop-on-computation-learnability-and-phonological-theory>)
- 2015 38th Generative Linguistics in the Old World (GLOW) colloquium; co-organizer; Paris, April 15-17 (<https://sites.google.com/site/2015glow/home>)
- 2014 Theoretical Issues in Contemporary Phonology: Reading Tobias Scheer; co-organizer; Paris, February 6-8 (<https://sites.google.com/site/readeingtobiasscheerenglish/>)
- 2012 *Cognitive Computational Phonology Workshop*; Paris, 10 November (<https://sites.google.com/site/ccpmeeting/>)
- 2007 *SNEWS: the Southern New England Workshop in Semantics*; MIT, October 20



**B.16. Academic service**

- Departmental service: • Co-organizer of the *Atelier de Phonologie* (regular phonology seminar at SFL UMR 7023); 2012-present (<https://sites.google.com/site/atelierphono/>)
- Organizer of the *SIGMA seminar* (regular linguistic seminar at the Institute Jean Nicod, ENS); 2010-2011
- Journal reviewer: **Natural Language and Linguistic Theory** (2005; 2007; 2009; 2012) | **Natural Language Semantics** (2005; 2006; 2008; 2010; 2011; 2014; 2015) | **Journal of Semantics** (2007; 2009; 2011; 2012; 2014) | **Linguistic Inquiry** (2008; 2008bis; 2009; 2013; 2014; 2015) | **Poznan Studies in Contemporary Linguistics** (2010) | **Mind and Language** (2007) | **Linguistics** (2009) | **Journal of Logic, Language and Information** (2012) | **The Linguistic Review** (2012) | **Phonology** (2012; 2013; 2014) | **Computational Linguistics** (2015).
- Conference reviewer: **ESSLLI** (European Summer School in Logic, Language, and Information: 2012) | **FDSL** (Formal Description of Slavic Languages Conference: 2014) | **Genius** (Paris Genericity Conference: 2010) | **GLOW** (Generative Linguistics in the Old World: 2010; 2015) | **NASSLLI** (North American Summer School in Logic, Language and Information: 2014) | **NELS** (North East Linguistic Society: 2014; 2015) | **AMP** (Annual Meeting in Phonology: 2014; 2015) | **RALFe** (Rencontres d'Automne de Linguistique formelle: 2012) | **SIGMORPHON** (ACL Special Interest Group in Morphology and Phonology: 2012; 2014) | **SuB** (Sinn und Bedeutung: 2012; 2015) | **MOL** (Mathematics of Language: 2015)
- Grants reviewer: Marie Curie Zukunftscolleg Incoming Fellowship Programme | The United States/Israel Binational Science Foundation
- Coeditor: *Proceedings of the MIT-France Workshop on Implicatures and Presuppositions*, with Paul Egré; MITWPL volume 60, 2009