

BIASES AND STAGES IN PHONOLOGICAL ACQUISITION

A Dissertation Presented

by

ANNE-MICHELLE TESSIER

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

DOCTOR OF PHILOSOPHY

February 2007

Linguistics

© Copyright by Anne-Michelle Tessier 2007

All Rights Reserved

BIASES AND STAGES IN PHONOLOGICAL ACQUISITION

A Dissertation Presented

by

ANNE-MICHELLE TESSIER

Approved as to style and content by:

Joseph V. Pater, Chair

John J. McCarthy, Member

Lyn Frazier, Member

Shelley L. Velleman, Member

Elizabeth O. Selkirk, Department Head
Linguistics

DEDICATION

To Elizabeth Tessier and Attilio Favro.

ACKNOWLEDGEMENTS

It has now become cliché to begin one's acknowledgements by acknowledging that they are the part most likely to be read in any linguistics dissertation. I apologize for the cliché but not for its truth; since I plan to continue reading other people's acknowledgements before all else, I can only offer the hope that readers enjoy these acknowledgements enough to read on until perhaps the table of contents.

First I want to express my gratitude to my committee members: Joe Pater, John McCarthy, Lyn Frazier and Shelley Velleman. My chair Joe Pater has guided me gently through every obstacle of graduate school; I thank him for helping me construct and then improve every piece of this dissertation and for always making the effort to figure out what I was trying to think or write and extracting the salvageable parts. I also owe him for involving me in his own work from the very beginning, and for teaching me about being a theoretician who runs experiments, a phonologist who talks to psychologists, and a linguist who still goes to rock concerts. I thank John McCarthy for being, well, John McCarthy: for applying the full power of his empirical and technical knowledge to all of my work, for reading and improving drafts nearly before I gave them to him, and for being so devastatingly clear. I also thank him for teaching the first UMass phonology seminar I attended, which convinced me I was doing the right thing, and for setting the bar for teaching phonology and Optimality Theory so instructively high. I thank Lyn Frazier for being a voice of reason, breadth and perspective, for her experimental knowledge, for always asking the trickiest questions, and for sharing her enthusiasm for connecting data and theory, making strong claims and thinking very hard. I thank Shelley

Velleman for providing her expertise on phonological acquisition and disorders, for carefully understanding my perspective and carefully reading my work, for pointing me to all the data I didn't know, and for smiling in the face of stress.

Beyond my committee, I thank John Kingston for teaching me most of the correct things that I know about doing and thinking about phonetics, and for always being willing to discuss any theory or experiment as clearly as I would let him, and I thank Lisa Selkirk for challenges and support throughout my time at UMass. Among UMass phonology students past and present, I thank Kathryn Flack, Michael Becker, Shigeto Kawahara, Tim Beechey, Matt Wolf, Maria Gouskova, Andries Coetzee, Mike Key, Karen Jesney, Kathryn Pruitt, Della Chambless, Jonah Katz and Dan Mash, for data, theory, advice, argumentation, proofreading, scripts, stats, commiseration, deep-fried food and all manner of collegial help. I am so very grateful to have worked in this department's unique, talented and supportive community of teachers and peers; I hereby acknowledge how much I owe you all.

Many other UMass people must also be thanked for getting me through grad school. I would like to thank Marcin Morzycki for everything, only I don't know how, so instead I will just thank him for arguing with me – about linguistics, pedagogy and nearly all else, at every hour of the day or night, to extents no doubt in violation of aspects of the Geneva convention – for lending me platypuses, and for making me laugh. With similar insufficiency, I thank Ana Arregui and Paula Menéndez-Benito, for every kind of love, support and sage advice that friends can give, for calling and emailing and checking and comforting and listening, and for making me laugh. I thank Jan Anderssen and Meredith Landman for being such wonderful roommates and sharing so many of the daily horrors

of graduate work and dissertation writing, as well as the food in the fridge, and for making me laugh. I thank my classmates Helen Majewski, Helen Stickney and Shai Cohen, for support both quiet and loud, mutual miseries and reality checks, and for making me laugh. And for many acts of friendship along the way, I thank Keir Moulton, Michael Brigham, Florian Schwartz, Ilaria Frana, Aynat Rubenstein, Kyle Rawlins, Amy Rose Deal, Masako Hirotani and Tessa Warren, as well as everyone else who ever came to Semantics Reading Group and drank my martinis, or made me laugh. I also thank Rajesh Bhatt for dissertation-stimulating music and party-hosting, Angelika Kratzer for being so sweetly happy for me when I got a job, Ellen Woolford for not giving up on the hope that I would someday say something clearly, and Kathy Adamczyk and Tom Maxfield for continually saving my administrative hide.

Outside UMass, I thank the following people whose time, insight, questions and critiques improved many parts of this dissertation: Bruce Hayes, Bruce Tesar, Elan Dresher, Keren Rice, Kie Zuraw, Colin Wilson, Adam Albright, Glyne Piggott, Heather Goad, Marie-Hélène Côte, Kevin Ryan, Jason Riggle, Elliott Moreton, Jennifer Smith, Michael Wagner, and audiences at the University of Toronto, the MOT, HUMDRUM, and Brown/UMass phonology workshops, the 80th LSA meeting, BUCLD30 and WCCFL25. I also thank Alan Prince for his LSA 2005 summer institute course, the Social Sciences and Humanities Research Council of Canada (a.k.a. Money Canada) for their generous support over the last two years, and the Department of Linguistics at the University of Alberta for giving me such a good reason to finish.

At McGill University, I would like to thank my teachers and colleagues who convinced me that I should be a linguist by treating me like one – including Lisa Travis,

Jonathan Bobaljik, Yvan Rose, Evan Mellander, Lydia White, Lara Riente, Charles Boberg and Ben Adaman. In particular I thank Heather Goad, for getting me hooked on phonology, OT and acquisition, sending me to UMass for brainwashing, and then setting an impeccable example of how to support and happily disagree with a former student.

I thank my parents Roger and Rosemary Tessier for always believing I could do whatever I thought I wanted to, for covering all my computational and laundry needs respectively, for their stellar genetic material, and for making me laugh. And I thank Allon Beck and Rebecca Rosenblum for being permanent friends, and for each coming to visit western Massachusetts when friendship at a distance was not enough.

Finally, I thank Kyle Johnson – for teaching me many valuable lessons not discussed in the graduate school handbook, for being charming even beyond the many times he picked up the cheque, for making me laugh, and for preventing me from quitting linguistics when I was at my lowest by pointing out I'd be just as terrible at anything else.

ABSTRACT

BIASES AND STAGES IN PHONOLOGICAL ACQUISITION

FEBRUARY 2007

ANNE-MICHELLE TESSIER, B.A., MCGILL UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Joseph V. Pater

This dissertation presents Error-Selective Learning, an error-driven model of phonological acquisition in Optimality Theory which is both *restrictive* and *gradual*. Together these two properties provide a model that can derive many attested intermediate stages in phonological development, and yet also explains how learners eventually converge on the target grammar.

Error-Selective Learning is restrictive because its ranking algorithm is a version of Biased Constraint Demotion (BCD: Prince and Tesar, 2004). BCD learners store their errors in a table called the Support, and use ranking biases to build the most restrictive ranking compatible with their Support. The version of BCD adopted here has three such biases: (i) one for high-ranking Markedness (Smolensky 1996) (ii) one for high-ranking OO-Faith constraints (McCarthy 1998; Hayes 2004); and (iii) one for ranking specific IO-Faith constraints above general ones (Smith 2000; Hayes 2004).

Error-Selective Learning is gradual because it uses a novel mechanism for introducing errors into the Support. As errors are made they are not immediately used to learn new rankings, but rather stored temporarily in an Error Cache. Learning via BCD is only triggered once some constraint has caused too many errors to be ignored. Once

learning is triggered, the learner chooses one *best* error in the Cache to add to the Support – an error that will cause minimal changes to the current grammar.

The first main chapter synthesizes the existing arguments for this BCD algorithm, and emphasizes the necessity of the Support's stored errors. The subsequent chapter presents Error-Selective Learning, using cross-linguistic examples of attested intermediate stages that can be accounted for in this approach. The third chapter compares ESL to a well-known alternative, the Gradual Learning Algorithm (GLA: Boersma, 1997; Boersma and Hayes, 2001), and argues that the GLA is overall not well-suited to learning restrictively because it does not store its errors, and because it cannot reason from errors to rankings as does the BCD. The final chapter presents an artificial language learning experiment, designed to test for high-ranking OO-faith in children's grammar, whose results are consistent with the biases and stages of Error-Selective Learning.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	v
ABSTRACT.....	ix
CHAPTER	
I INTRODUCTION.....	1
1. Summary of the dissertation	1
2. Structure of the dissertation	3
II PHONOTACTIC LEARNING AND BIASED CONSTRAINT DEMOTION.....	6
1. Introduction.....	6
1.1 Important aspects of Optimality Theory for this learning theory.....	7
1.2 Outline of the chapter.....	8
2. Learning an Optimality-Theoretic Grammar	9
2.1 The learning framework: the Tesar/Smolensky learner.....	9
2.2 The importance of errors and the Support.....	13
2.2.1 The Support and the lexicon.....	14
2.2.2 The Support and the Gradual Learning Algorithm.....	15
3. Restrictive phonotactic learning: Biased Constraint Demotion	16
3.1 Illustrating the BCD approach: high-ranking Markedness	19
3.2 A second bias in BCD terms: high-ranking OO-Faith	23
3.2.1 OO faith as an OT account of cyclicity	24
3.2.2 OO-faith as an OT account of MSCs	26
3.2.3 The learnability argument for high-ranking OO-faith: McCarthy (1998).....	28
3.3 Connecting M >> F and OO >> IO as surface-oriented biases.....	29
3.4 Interim Summary	33
4. An input-oriented ranking bias in BCD: Specific-F >> General-F	33
4.1 The theory of positional faithfulness	34

4.1.1 Why not (only) positional markedness.....	35
4.1.2 Stringency, not fixed rankings.....	42
4.2 The learnability argument for a Specific-F >> General-F bias: Smith (2000).....	43
4.3 The problems of enforcing the Specific-F >> General-F bias.....	46
4.3.1 Language-specific relations between faithfulness constraints.....	46
4.3.1.1 Prince and Tesar's example	46
4.3.1.2 A morphological example	48
4.4 Interim Summary.....	49
5. Returning to the role of the Support	50
5.1 A kind of learning error: winner misparses	51
5.2 How the Support allows BCD to overcome winner misparses.....	53
5.3 A second example	55
5.4 Summary.....	57
6. The proposal: finding the most specific IO-Faith constraint.....	58
6.1 The goal: determining subset relations between the contexts of faith.....	58
6.1.1 Constraint stringency vs. context specificity.....	59
6.1.2 Outline of the proposal.....	60
6.2 The first step: finding universal specificity relations	63
6.3 The second step: finding contingent specificity relations	66
6.4 Why context tables are dynamic	70
6.4.1 What can go wrong in a context table?	71
6.4.2 Overcoming extra marks in a context table	74
6.5 Why contingent specificity cannot be learned from Ls and Ws	75
7. Implementing the Spec-F >> Gen-F bias	77
7.1 A working BCD algorithm	77
7.2 Ways of calculating Spec-F >> Gen-F relations: the Azba case study....	79
7.2.1 Using a context-based F-specificity bias	81
7.2.2 Using constraint-based F-stringency bias: Hayes' simulation	83
7.2.3 Prince and Tesar on the Azba language	85
7.2.4 Summarizing the Azba results	86

7.3	Returning to Anti-Paninian rankings and learning	87
7.4	Summary and outstanding issues	90
8.	Chapter 2 Summary, in preparation for Chapter 3	91
III	ERROR-SELECTIVE LEARNING.....	94
1.	Introduction.....	94
1.1	The approach to reconciling BCD and <i>gradual</i> learning	95
1.2	The Specific Markedness stage: English coda clusters	98
1.3	The Specific Faithfulness stage: French onset clusters	99
1.4	Analytic assumptions about the intermediate stages.....	101
1.4.1	Stringency relations among markedness constraints.....	101
1.4.2	Positional faithfulness and input prosodic structure.....	102
1.5	Roadmap to the chapter.....	105
2.	The data from intermediate stages	105
2.1	Introduction to the data	105
2.1.1	The Compton/Streeter database	106
2.2	Intermediate stages that rely on specific markedness	107
2.2.1	More on complex codas in Germanic.....	107
2.2.2	Markedness of complex onsets, and sonority distance	108
2.3	Intermediate stages that rely on specific faithfulness	115
2.3.1	More on faithfulness in stressed syllables	115
2.3.2	Faithfulness <i>to</i> stressed syllables.....	120
2.3.3	Faithfulness to initial syllables.....	121
2.3.4	Faithfulness to morphological roots	122
2.4	Summary of the data	125
3.	The theory of intermediate stages: Error-Selective BCD	126
3.1	The Error-Selective Learning proposal	126
3.1.2	What happens when an error is made: a Specific-M example....	127
3.1.2	How learning is triggered	128
3.1.3	Step 1: Choosing an error to learn from.....	129
3.1.4	Step 2: Applying BCD.....	130
3.1.5	A second example: a Specific-F stage.....	131

3.2	Discussion of the ESA, and Error-Selective Learning more generally	133
3.2.1	Analyzing the three ESA criteria for choosing errors.....	133
3.2.2	Terminating ESL and converging on the end stage grammar.....	135
3.2.3	Irrelevant markedness violations	137
3.2.4	Choosing among positional faithfulness contexts.....	138
3.2.5	The Violation Threshold and extra-grammatical factors.....	139
3.3	Illustrating ESL: A case study of Trevor and Julia's onset clusters.....	140
3.3.1	Trevor	140
3.3.2	Julia	147
3.3.3	Summary	151
4.	The roles of frequency	152
4.1	The connection between frequency and Error-Selective Learning	152
4.2	The connection between frequency and order of acquisition	153
4.2.1	Data from cross-linguistic frequency: initial weak syllables vs. codas	154
4.2.2	Ambient not output frequencies, and the Error Cache.....	156
4.3	Intermediate stages without stringency: stages of prosodic truncation ..	160
4.3.1	Noting a stringent alternative	165
4.4	Infrequent mistakes and the value of the Error Cache	166
5.	Developmental variation and Error Selective Learning	168
5.1	The ubiquity and challenges of variation in learning	169
5.1.1	The potential for a variable BCD learner.....	170
5.2	Alternative I: the Variable VT approach.....	172
5.2.1	The example of variable codas.....	173
5.2.2	The effects of the variable VT approach.....	177
5.2.3	Deriving developmental regression in the Variable VT approach.....	179
5.2.4	Weaknesses of the Variable VT approach.....	180
5.3	Alternative II: the Cloned Support approach	180

5.3.1	Returning to the variable coda example.....	181
5.3.2	Discussion of the Cloned Support approach.....	184
5.3.3	Regression in the Cloned Support approach.....	185
5.4	Summarizing the variable ESL discussion.....	186
6.	Chapter Summary	187
IV.	THE GRADUAL LEARNING ALGORITHM ALTERNATIVE.....	188
1.	An introduction to the Gradual Learning Algorithm (GLA).....	188
1.1	The GLA view of constraint rankings.....	188
1.2	How the GLA learns a grammar.....	191
1.2.1	The (limited) power of an error in the GLA	193
1.3	Goals and core properties of the GLA	196
1.4	Ranking Biases and the GLA.....	197
1.5	Chapter Roadmap.....	198
2.	Restrictiveness and specific-to-general faithfulness relations in the GLA.....	198
2.1	The exemplifying grammar	199
2.2	The GLA's learning input	200
2.3	The stages of GLA learning	203
2.3.1	The initial state	203
2.3.2	The intermediate stages	204
2.3.3	The end state grammar	206
2.4	Summarizing the results	207
2.4.1	The superset grammar: mid vowels	207
2.4.2	The Anti-Paninian Ranking: round vowels	208
3.	Intermediate stages and the Specific-F >> General-F bias in the GLA.....	209
3.1	The Specific F stages that require the ranking bias	209
3.2	The Specific F stages that don't require the bias: Curtin and Zuraw (2001)	211
3.3	Interim Summary	214

4.	Persistent biases, contingent biases, and the GLA	214
5.	A first problem with not storing errors: winner misparses.....	217
5.1	The GLA's treatment of misparsed winners	218
5.2	Winner misparses and markedness: the same problem	223
6.	Exceptions and end-state variation.....	225
6.1	The GLA's treatment of exceptionality.....	226
6.1.1	Two languages and their codas.....	227
6.1.2	Learning the variable coda grammar.....	228
6.1.3	Learning the exceptional coda grammar.....	230
6.2	Learning exceptions: GLA-related approaches.....	232
6.3	Learning variation without the GLA: a BCD approach.....	236
6.4	Summary.....	243
7.	Chapter Summary.....	243
V.	TESTING FOR THE HIGH-RANKING OO-FAITH BIAS.....	245
1.	Introduction to the chapter	245
2.	The OO-faith ranking bias and phonotactic learning	246
2.1	The role of OO-faith in enforcing restrictiveness	246
2.2.	Predictions for stages of acquisition	249
2.2.1	The target: an OO-unfaithful language	249
2.2.2	OO-faith kicks in at the initial state	250
2.2.3	OO-faith kicks in at an intermediate state	251
3.	The experimental methodology: artificial language learning	253
3.1.	The difficulties in testing for OO-faith in L1 acquisition	253
3.2.	The artificial language learning paradigm	254
3.3	The present application	255
4.	Experimental Design	257
4.1	Experimental predictions	257
4.2	Materials	258
4.3	Methodology	259

4.3.1	Participants	260
4.3.2	Training	260
4.3.3	Testing	262
5.	Experimental Results	263
5.1	The data reported	263
5.2	Testing the predictions	264
5.2.1	Testing prediction 1	264
5.2.2	Testing prediction 2	266
5.3	Summary of the results	267
6.	Rankings in the results	267
6.1	A3's cluster voicing: an intermediate ranking	268
6.2	A3's treatment of coda affricates	269
6.3	N's treatment of [mf.d] clusters: an initial state ranking	271
6.4	Summary of analyses	273
7.	Theoretical discussion	273
7.1	The intermediate stage, and Error-Selective Learning.....	273
7.2	Independent evidence of the OO-faith bias	275
7.3	The persistent OO-faith bias, and the GLA	280
7.3.1	The empirical need for a persistent OO-faith bias	280
7.3.2	The GLA problem with persistent biases and OO-faith	281
8.	Experimental discussion	284
8.1	The connection between natural and artificial language learning	284
8.2	A potential perceptual confound, and the next step	286
9.	Chapter Summary.....	287
	BIBLIOGRAPHY.....	288