Probabilistic Selection of Input in Morphophonological Acquisition

by

Tsung-Ying Chen

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics
University of Alberta

Abstract

This dissertation sets out to explain the development during morphophonological acquisition and its possible learning outcomes by constructing a Probabilistic Selection of Input (PSI) rich lexicon learning model, in part based on psycholinguistic evidence that rich language details are lexically encoded. Contrary to traditional UR assumptions based on lexical economy, PSI stores and associates all surface allomorphs of a morpheme in a rich lexicon as possible inputs of the morpheme. Through the lexical associations between stored allomorphs, the leaerner assigns a probability between 0 and 1 to each allomorph and probabilistically select the phonological input of the morpheme. Output pattern variation along the acquisition course is thus analyzed as results of different input preferences and corresponding grammar shifts at sequential stages, and a successful morphophonological learning stands for an adult-like lexical generalization (i.e. input probabilities) and phonological grammar captured by learners. Diachronic morphophonemic changes can nevertheless occur with a shift in input preferences over learning generations, which gradually leads to permanent grammar shifts.

PSI is tested with computer simulations using corpus data as a training corpus in various case studies, including the acquisition of Dutch stem-final voicing alternation (Chapter 3), the diachronic change of Mandarin Tone 3 (Chapter 4), and

the emergence of Korean stem-final obstruent variations (Chapter 5). Learning outcomes similar to the performance by native speakers in elicitation tasks are demonstrated in the PSI simulations as a result of temporarily or permanently selecting different stored allomorphs as phonological inputs.

This thesis is dedicated to the eternal memory of my father Chin-Sui Chen.

Acknowledgements

There are numerous people who have greatly contributed to the completion of my degree. I would like to take this opportunity to truly express my gratitude to my colleagues, friends, and family.

My sincere gratitude firstly goes to my most respected supervisors Robert Kirchner and Anne-Michelle Tessier, who have always been supportive and encouraging and generously shared their profound knowledge to guide me through the five years of my study at Edmonton. I am also greatly benefited from the collaboration with a member of my supervisory committee, Benjamin V. Tucker, and always enjoyed our conversation concerning linguistics. Other faculty members, including Harald Baayen, Patrick Bolger, Terry Nearey, and Sally Rice, and administrative staff members have all played different roles during my exploration to the human mind at the University of Alberta, and also have made my PhD journey much smoother; for this I appreciate it very much.

I am also grateful to the following linguists outside the department for their kind assistance: Greg Kondrak and Gaja Jarosz provided very insightful suggestions and comments from different perspectives during my candidacy and thesis defense, which have also made my thesis a work of quality. As part of my connection to the linguistic circle in Taiwan, James Myers and Hui-Chuan Hsu have been especially encouraging.

We need friends to create a more entertaining PhD life for sure. The following are those who have been the best among them. Corey Telfer has always been excellent company when I want to talk about linguistics and ice hockey (or basically whatever you prefer). Dagmara Dowber is a very thoughtful friend to whom I personally owe a lot in various aspects. There are a few more from the department on my fun list: Simon Fung, Doug Inglis, Ruiting Jia, Caelan Marr, Kevin Penner, Hien Pham, Vincent Porretta, Rebbeka Puderbaugh, Michelle Sims, and Guangting Wang. Also, my special thanks to Chao-Ting Chou and Cheng-Hung Hsin for our everlasting friendships.

Last but not least, I am forever indebted to my family – my mother Yin-Fen, my two sisters Chiu-Hui and Ying-Ling – for their understanding to my personal pursuit and long absence. I have the same words to my beloved significant other, Wen-Chen, for her love and patience.

Table of Contents

List of Tables xii

List of Tableaux xvi

List of Figures xxi

List of Acronyms xxvi

Chapter 1. Introduction 1

1. Lexical abstractness 2
2. Circularity problem 6
3. Relexicalization and rule inversion 7
4. Evidence for a rich lexicon 9
   - 4.1. Against abstractness: Phonetic details in child and adult phonology. 9
   - 4.2. Against parsimony I: Lexical entries and token frequency 12
   - 4.3. Against parsimony II: Multidirectional paradigmatic leveling 15
5. Rich lexicon and phonological generalization 17
6. Probabilistic Selection of Input: Introduction 19
7. Probabilistic Selection of Input and constraint grammar 21
8. Design guidelines of a learning algorithm 21
9. Summary 23

Chapter 2. Probabilistic Selection of Input with constraint grammar 25

1. Selection probability of lexically stored allomorphs 25
   - 1.1. Lexical factor I: Token frequency 25
   - 1.2. Lexical factor II: Individual Error Proportion 27
   - 1.3. Lexical factor III: Morphological Error Proportion 32
   - 1.4. Lexical factor IV: Memory decay 40
   - 1.5. Summary of Probabilistic Selection of Input 41

2. Constraint grammar and Gradual Learning Algorithm 42
2.1. Innate and learnable constraints 42
2.2. Initial state of constraints and re-ranking process 47
2.3. Gradual changes and random noise in constraint values 49
2.4. Gradual Learning Algorithm: Caveats 52
3. Two-stage (morpho-)phonological acquisition 53
4. Summary of PSI-OT-GLA learning process at different stages 56
5. Testing PSI-OT-GLA with toy data 58
5.1. Training corpus and input distribution 59
5.2. Constraint set, initial state, and target grammar 59
5.3. Results: Grammar learning 60
5.4. Results: Lexical learning 63
5.5. Local summary 65
6. Advantages over previous morphophonological learning models 67
6.1. Inconsistency detection 67
6.2. Maximum Likelihood Learning 68
6.3. Lexical constraint 69
6.4. Summary 71

Chapter 3. Case study I: The acquisition of Dutch final devoicing 73
1. Children's production of Dutch voicing alternation 73
2. Training corpus and constraint set 76
2.1. Input distribution 77
2.2. Structure of inputs and correct outputs 79
2.3. Constraint set 80
3. Elicitation task in computer simulation 82
4. Simulated results 83

4.1. Phonotactic learning ... 83
4.2. Morphophonological learning: Prediction ... 86
4.3. Grammar learning after morphological awareness ... 89
4.4. Lexical learning after morphological awareness ... 92
4.5. Simulated results vs. experiment results ... 98
4.6. Discussion ... 101
5. Testing alternative models with fewer lexical factors ... 104
5.1. Naïve Parameter Learning without lexical factors ... 104
5.2. Frequency plus Individual Error Proportion ... 108
5.3. Frequency plus Morphological Error Proportion ... 109
5.4. PSI-OT-GLA without memory decay ... 110
5.5. Summary of model comparison ... 114
6. Local summary ... 115

Chapter 4. Case study II: Tone 3 change in Standard Mandarin ... 116
1. Tone sandhi in Standard Mandarin ... 117
2. Unexceptional Tone 3 reduction in Taiwan Mandarin ... 124
2.1. Chain-shift tonal change in Taiwan Mandarin ... 124
2.2. A PSI-based account ... 126
3. Testing Tone 3 production in Standard Mandarin ... 132
3.1. Materials ... 133
3.2. Procedure ... 135
3.3. Participants ... 137
3.4. Result ... 138
3.5. Analysis ... 141
3.6. Discussion ... 145
4. Modeling Tone 3 change with PSI-OT-GLA ... 146

4.1. Input structure 146
4.2. Constraint set, output candidates, and target outputs 147
4.3. Lexical factors 151
4.4. Perception and production in two-stage learning 152
4.5. Learning over generations 153
4.6. Elicitation task in computer simulation 153
4.7. Results 154
4.8. Discussion 160
5. Local summary 162

Chapter 5. Case study III: Stem-final obstruent variations in Korean 163

1. Standard Korean phonology 163
1.1. Obstruent coda neutralization in bare forms 164
1.2. Assibilatory affrication 168
1.3. Summary: Constraint ranking of standard pronunications 171
2. Stem-final obstruent variations in Korean suffixed noun forms 171
3. A PSI-based account 175
3.1. Against stem-final plain /t/ in suffixed nouns 176
3.2. Against stem-final non-plain /p/ and /k/ in suffixed nouns 180
3.3. Locative/dative (and other non-bare) allomorphs as stem inputs .. 181
3.4. After the emergence of stem-final variations 184
3.5. Summary of the stem-final variation development 186
4. Modeling Korean stem-final variations with PSI-OT-GLA 188
4.1. Training corpus, input distribution, and input structure 188
4.2. Constraint set and initial state 192
4.3. Lexical factors 194
4.4. Perception and production in two-stage learning 194

4.5. Learning over generations and a simulated elicitation task 196
4.6. Results 196
4.6.1. Grammar devlopment for stem-final coronal variations 197
4.6.2. Grammar devlopment for stem-final labial/velar variations 200
4.6.3. Development of Morphological Error Proportion 201
4.6.4. Similarities between simulated and experiment results 203
4.6.4.1 Stem: 'sickle' 203
4.6.4.2 Stem: 'day' 204
4.6.4.3 Stem: 'flower' 206
4.6.4.4 Stem: 'wall' and 'rice' 207
4.6.5. Mismatches between simulation and experiment result 208
4.6.5.1 Stem: 'clothes' 208
4.6.5.2 Stem: 'breast' 210
4.6.5.3 Stem: 'face' 211
4.6.5.4 Stem: 'field' 212
4.6.5.5 Stem: 'red bean' 213
4.6.5.6 Stem: 'kitchen', 'outside', and 'leaf' 215
4.7. Discussion 217
4.8. Advantages over base identification approach 218
5. Local summary 221

Chapter 6: Toward an Output-Output correspondence model 222
1. Success and failure of PSI-OT-GLA 223
2. A rich lexicon and Output-Output correspondence 225
2.1. Lexical entry, lexical paradigm, and lexical associations 226
2.2. Lexically-indexed Output-Output correspondence 228
2.3. LIOO in Harmonic Grammar 231

2.4. Initial state of constraints and weight adjustment in LIOO-HG.....233
2.5. Semantic similarity and OO correspondence in LIOO-HG...........235
3. Potential improvements in previous simulations with LIOO-HG.........237
3.1. Dutch final devoicing.........238
3.2. Mandarin Tone 3 reduction.........244
3.3. Korean stem-final variations.........245
3.4. Local summary.........248
4. Additional advantages without input selection in LIOO-HG.........248
5. Closing remarks.........251

Bibliography.........253
Appendix.........272

List of Tables

Table 2.1. Input distributions with three frequency ratios of [bap]:[ba.b-a]...........59

Table 2.2. Constraint values/rankings acquired with three bias ratios in the P stage simulation...61

Table 2.3. Constraint values/rankings acquired at the 10,000th cycle in the M stage simulation...63

Table 3.1. Non-alternating (NAP) and alternating (AP) Dutch plural forms with their target output and token frequency...74

Table 3.2. Output type rates and token numbers of alternating plural forms (AP) and non-alternating plural forms (NAP)...75

Table 3.3. Devoicing error rates and total error tokens of individual alternating plural forms in the elicitation task...76

Table 3.4. Token frequencies, distributional probabilities, and singular-plural ratios of the 32 Dutch forms from CELEX...78

Table 3.5. Constraint values/rankings acquired at the end of the P stage simulation...83

Table 3.6. Constraint values/rankings acquired at the end of the P stage simulation...91

Table 3.7. Constraint values/rankings acquired at the end of the M stage simulation (i.e. the 100,000 cycle)..92

Table 3.8. The SP of plural allomorphs at the end of the M stage simulation...........93

Table 3.9. Comparison of plural form error rates between simulated and real experiments...99

Table 3.10. Devoicing error rates in Kerkhoff's experiment and simulation.........100

Table 3.11. Constraint values/rankings acquired at the end of the M stage simulation in NPL-OT-GLA..106

Table 3.12. The SP of plural allomorphs at the end of the M stage simulation in NPL-OT-GLA......106
Table 3.13. Comparison of plural output error rates between simulated and real experiments......107
Table 3.14. Constraint values/rankings acquired at the end of the M stage simulation in the Freq+IEP model......109
Table 3.15. The SP of plural allomorphs at the end of the M stage simulation in the Freq+IEP model......109
Table 3.16. Constraint values/rankings acquired at the end of the M stage simulation in the Freq+MEP model......110
Table 3.17. The SP of plural allomorphs at the end of the M stage simulation in the Freq+MEP model......110
Table 3.18. Constraint values/rankings acquired at the end of the M stage simulation in the PSI-OT-GLA without memory decay......111
Table 3.19. The SP of plural allomorphs at the end of the M stage simulation in PSI-OT-GLA without memory decay......112
Table 3.20. Comparison of plural output error rates between simulated and real experiments......113
Table 4.1. Contextual token frequency of Tone 3 words selected for the elicitation experiment......134
Table 4.2. By-item proportions of the full concave variant in phrase-final positions (MLH proportion) in the second task......141
Table 4.3. Linear Mixed-Effect Regression analysis of the Tone 3 word production......143
Table 4.4. Linear Mixed-Effect Regression analysis of the Tone 3 word production with the additional predictor Dialect......144
Table 4.5. Linear Mixed-Effect Regression analysis of the Tone 3 word production

of the seven speakers showing variations in the third task....144
Table 4.6. Frequency proportions (i.e. distributional probabilities for Tone 3 word selection) of the 18 target Tone 3 words in different contexts....147
Table 4.7. P-values of the sole independent variable Ratio in linear regression tests using concave percentages as the dependent variable....159
Table 5.1. Standard pronunciation of thirteen noun stems in different morphosyntactic contexts....164
Table 5.2. Token numbers of stem-final coronal obstruents used in suffixed noun forms (Jun & Lee 2007, Experiment I)....173
Table 5.3. Stem-final bilabial/velar obstruent variations in Jun & Lee's Experiment I....174
Table 5.4. Frequency and ranking of possible extensions....180
Table 5.5. Token frequency of thirteen Korean noun stems in five morphosyntactic contexts....189
Table 5.6. Corrected token frequency and distributional probability (parenthesized) of thirteen Korean noun stems in five morphosyntactic contexts....190
Table 5.7. An example of lexical factor matrix of the stem 'field' and the possible surface allomorphs in its standard pronunciations....195
Table 5.8. Percentages of the primary and secondary outputs of 'sickle' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)....203
Table 5.9. Percentages of the primary and secondary outputs of 'day' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)..205
Table 5.10. Percentages of the primary and secondary outputs of 'flower' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)....206

Table 5.11. Percentages of the primary and secondary outputs of 'wall' and 'rice' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)....208
Table 5.12. Percentages of the primary and secondary outputs of 'clothes' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs....208
Table 5.13. Percentages of the primary and secondary outputs of 'breast' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)....210
Table 5.14. Percentages of the primary and secondary outputs of 'face' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)....211
Table 5.15. Percentages of the primary and secondary outputs of 'field' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)....213
Table 5.16. Percentages of the primary and secondary outputs of 'red bean' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)....214
Table 5.17. Percentages of the primary and secondary outputs of 'kitchen' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs....215
Table 5.18. Percentages of the primary and secondary outputs of 'outside' and 'leaf' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs....216
Table 6.1. Mini-corpus with raw token frequencies in CELEX (Baayen et al. 1995) and corresponding distributional probabilities....238

List of Tableaux

Tableau 2.1. Demotion of markedness constraint and promotion of faithfulness constraint....48

Tableau 2.2. Production of correct output after constraint re-ranking....48

Tableau 2.3. Coda deletion error with the rankings acquired in the P stage....62

Tableau 2.4. Constraint re-ranking triggered by an input-output mapping error..70

Tableau 2.5. No effect of lexical constraints with dominant markedness and faithfulness constraints....71

Tableau 3.1. Promotion of IDENT(voi)-IO to preserve input voiced obstruents.....84

Tableau 3.2. Promotion of IDENT(voi)-IO to outrank *V[-voi]V....84

Tableau 3.3. Promotion of DEP-IO to prohibit the insertion repair to marked syllable structures....84

Tableau 3.4. MAX-IO is promoted over *VOICEDOBS to preserve voiced obstruents....85

Tableau 3.5. MAX-IO is promoted over *V[-voi]V and *CODA to avoid deletion..85

Tableau 3.6. Deleting an underlying voiced obstruent coda with the P stage ranking....86

Tableau 3.7. Faithful production of /bɛt/....87

Tableau 3.8. Output error from the input /bɛt+ən/....87

Tableau 3.9. Deleting an underlying voiced obstruent coda with the P stage ranking....88

Tableau 3.10. Output errors of non-alternating plural forms caused by optional intervocalic voicing....89

Tableau 3.11. Devoicing voiced obstruent coda after promoting DEP-IO and MAX-IO....89

Tableau 3.12. Preserving the intervocalic voicing contrast in Dutch with OO-F.102

Tableau 4.1. Tone 3 Sandhi triggered by top-ranked OCP-MLH....120
Tableau 4.2. Half Sandhi triggered by top-ranked *NONFINAL-MLH....120
Tableau 4.3. Phrase-final MLH is preserved with lower-ranked *MLH....121
Tableau 4.4. Other Tone 3 Sandhi outputs ruled out by higher-ranked *NONFINAL-MLH and MAX-LINK....121
Tableau 4.5. Output ML-MLH prohibited with a top-ranked *LONGLAPSE....122
Tableau 4.6. Selecting the optimal output with a shorter perceptual distance from the input....122
Tableau 4.7. Tone 2 MH prohibited by MAX-LINK(L/M_H) as the output of Half Sandhi....123
Tableau 4.8. Reduction of /MLH/ in non-final and phrase-final positions with top-ranked *MLH....126
Tableau 4.9. Neutralization between the inputs /ML/ and /MLH/ in the Half Sandhi context....127
Tableau 4.10. From the input MLH to the target output MH in the Tone 3 Sandhi context....128
Tableau 4.11. From the input ML to the target output MH in the Tone 3 Sandhi context....128
Tableau 4.12. Input MH surfaces faithfully in the Tone 3 Sandhi context....128
Tableau 4.13. Target outputs harmonically bounded by faithful outputs with the inputs /ML/ and /MH/....129
Tableau 4.14. /MLH/→[ML] with the ranking *MLH » MAX-LINK....130
Tableau 4.15. Limited types of output candidates of the input /H-MLH-MH-MLH/ with the initial ranking....150
Tableau 4.16. Surface errors generated from an underlying MH allomorph in the HS context....150

Tableau 4.17. The ranking *MLH » MAX-LINK required by the mapping MLH→ML....156

Tableau 5.1. Coda contrast neutralization in Korean....167

Tableau 5.2. Onset contrast preservation in Korean....168

Tableau 5.3. OT analysis of Korean assibilatory affrication....169

Tableau 5.4. Preserving affricates with a top-ranking IDENT(cont)-IO....170

Tableau 5.5. Non-derived environment blocking of assibilatory affrication....170

Tableau 5.6. OT analysis of assibilatory affrication with the sub-ranking *Ti » IDENT(strid)-IO....171

Tableau 5.7. Preference toward the faithful output [o.t-ɨl] with the input /ot+ɨl/.176

Tableau 5.8. Stem-final /t/ in suffixed nouns blocked by top-ranked *t/_]$_{STEM}$-S.177

Tableau 5.9. Conflicting constraint re-ranking of faithfulness constraints for either a /t/→[s] mapping or a /t/→[tʃ] mapping....178

Tableau 5.10. Extension of /t/→[s] mapping to 'breast (acc.)' and 'flower (acc.)'...179

Tableau 5.11. Constraint re-ranking of faithfulness constraints for a /t/→[t^h] mapping....179

Tableau 5.12. [$tʃ^h$] harmonically-bounded by [tʃ] with an underlying stem-final /t/....180

Tableau 5.13. Surface variations generated from bare form allomorphs in the inputs....181

Tableau 5.14. Standard pronunciations generated from other allomorphs in the inputs....181

Tableau 5.15. De-aspiration of stem-final labial and velar stops....182

Tableau 5.16. Surface variant [pa.$tʃ^h$-ɨl] from the input /pat^h+ɨl/ with higher-ranked *Tɨ....184

Tableau 5.17. Surface variants from the input /pat+ɨl/ with lower-ranked IDENT(cont)....185

Tableau 5.18. Surface variants from the input /pat^h+ɨl/ with higher-ranked *Tɨ..185

Tableau 5.19. Surface variants from the input /pat^h+i/ with higher-ranked *Ti...186

Tableau 5.20. Promotion of IDENT(asp)-IO against [Dor]/_]$_{STEM}$=[k]..................200

Tableau 5.21. Generating stem-final [tʃ^h] variants via OO Correspondence........218

Tableau 6.1. Paradigm leveling via input selection with higher ranked faithfulness constraints..225

Tableau 6.2. Violation of a LIOO constraint indexed to the paradigm [bɛd-ən]...229

Tableau 6.3. Paradigmatic leveling due to a higher frequency of [bɛt]..................230

Tableau 6.4. Faithful output due to a higher frequency of [bɛd-ən]......................230

Tableau 6.5. Paradigmatic leveling of singular forms is prohibited due to top-ranked *VOICEDOBSCODA...230

Tableau 6.6. Paradigmatic leveling of plural forms is prohibited due to higher-ranked *V[-voi]V...230

Tableau 6.7. Token frequency differences obscured by strict domination............231

Tableau 6.8. Token frequency bias capture in HG with a gang effect.....................233

Tableau 6.9. Weight adjustment triggered by output errors in HG.......................234

Tableau 6.10. Constraint promotion and demotion triggered by output errors in HG...235

Tableau 6.11. Semantic similarity between outputs and lexical paradigms as leveling pressure...237

Tableau 6.12. No violation of other LIOO constraints in the P stage; '✓' = correct output...239

Tableau 6.13. Violation(s) of LIOO constraints indexed to morphologically related paradigms in the M stage...239

Tableau 6.14. The weight of constraints crucial to final devoicing........................240

Tableau 6.15. Intervocalic voicing blocked by gang effect.....................................241

Tableau 6.16. Output error *[bɛ.t-ən] emerges with a higher output score;

constraint weights and accumulated token frequencies in the 100th learning cycle of the M stage (i.e. 10,100th learning cycle from the beginning of the simulation)....242

Tableau 6.17. Low score of output error *[vud-ən] with constraint weights and accumulated token frequencies in the 1,100th learning cycle of the M stage simulation....243

Tableau 6.18. Dominant correct output pattern [bɛd-ən] with constraint weights and token frequencies in the 10,000th learning cycle of the M stage simulation....243

Tableau 6.19. Paradigmatic leveling toward the half-tone verb paradigm....244

Tableau 6.20. Paradigmatic leveling toward the full concave noun paradigm....245

Tableau 6.21. Preserving stem-final segments in locative/dative context in Korean....246

Tableau 6.22. Extended [t]~[s] alternation in accusative context....247

Tableau 6.23. Extended [t]~[tʃʰ] alternation in an accusative paradigm....248

Tableau 6.24. Allomorph listing in phonologically conditioned allomorph selection....251

List of Figures

Figure 2.1. A global influence from MEP when producing a correct plural form with a plural allomorph.........37

Figure 2.2. Learning cycles in the P stage in PSI-OT-GLA; shaded numbers represent updated lexical information.........57

Figure 2.3. Learning cycles in the M stage in PSI-OT-GLA; shaded numbers represent updated lexical information.........58

Figure 2.4. Development of $SP_{/bab/}$, $IEP_{/bab/}$, and $MEP[X+a]_{PL}$ with the frequency ratio 1:1.........64

Figure 2.5. Development of $SP_{/bab/}$, $IEP_{/bab/}$, and $MEP[X+a]_{PL}$ with the frequency ratio 9:1.........65

Figure 2.6. Development of $SP_{/bab/}$, $IEP_{/bab/}$, and $MEP[X+a]_{PL}$ with the frequency ratio 19:1.........65

Figure 3.1. Development of constraint values in the initial 2,000 learning cycles in the M stage simulation.........91

Figure 3.2. Development of constraint values from the 2,100th to the 25,000th learning cycles in the M stage simulation.........92

Figure 3.3. Development of the SP (black solid line) and IEP (grey dotted line) of the plural allomorph of 'dog' before the 25,000th learning cycle.........93

Figure 3.4. Development of the SP (black solid line) and IEP (grey dotted line) of the plural allomorph of 'bed' before the 25,000th learning cycle.........94

Figure 3.5. Development of the $MEP[X+Ø]_{SG}$ (black solid line) and $MEP[X+ən]_{PL}$ (grey dotted line) from the 1st to the 10,000th learning cycle in the M stage simulation.........96

Figure 3.6. Development of the SP (black solid line) and IEP (grey dotted line) of the plural allomorph of 'hat' before the 25,000th learning cycle.........97

Figure 3.7. Changes in error rates from the 50th learning cycle to the 60,000th learning cycle....98
Figure 3.8. Comparison of alternating plural error rates between simulated and experimental results....100
Figure 3.9. Changes in error rates from the 50th learning cycle to the 35,000th learning cycle....107
Figure 3.10. Comparison of alternating plural error rates between NPL-OT-GLA, PSI-OT-GLA, and experimental results....108
Figure 3.11. Changes in error rates from the 50th learning cycle to the 6,000th learning cycle....112
Figure 3.12. Comparison of alternating plural error rates between PSI-OT-GLA, PSI-OT-GLA with memory decay, and experimental results....114
Figure 4.1. Production of [ma^{55}], [ma^{24}], [ma^{213}], and [ma^{51}] by a male native speaker of Standard Mandarin....117
Figure 4.2. Tonal representation of Tone 2 and Tone 3 in Standard Mandarin....119
Figure 4.3. Constraint ranking for Tone 3 Sandhi and Half Sandhi....124
Figure 4.4. Production of [pa] with Tone 2 (T2) and Tone 3 (T3) by a female speaker of Taiwan Mandarin in a carrier sentence *zhege zi shi nian* 'This character is read as ____'....125
Figure 4.5. A likely chain shift of Tone 2 and Tone 3 in Taiwan Mandarin....125
Figure 4.6. Constraint ranking after tonal changes in Taiwan Mandarin....126
Figure 4.7. (a) MLH of Tone 3 words represented by modal-creaky-modal sequence and (b) ML of Tone 3 words represented by modal-creaky sequence....139
Figure 4.8. (a) MLH of Tone 3 words represented by a concave amplitude contour and (b) ML of Tone 3 words represented by a falling amplitude contour....139
Figure 4.9. Predicted decreasing tendency in the proportion of the full concave

variants..143
Figure 4.10. Constructing a two-word phrase from selecting the lexical tone of each word to selecting the input of a Tone 3 word...147
Figure 4.11. Word-specific evaluation of output errors...152
Figure 4.12. Constraint values at the end of the M stage over 10 learning generations..155
Figure 4.13. SP_{MLH} of the six selected Tone 3 words over 10 learning generations..157
Figure 4.14. MLH proportion of the six selected Tone 3 words in the simulated elicitation task over ten learning generations...158
Figure 4.15. Comparison of concave percentages of the 12 Tone 3 words from the High and Medium ratio family between the simulated and experimental results..160
Figure 4.16. Hypothetical lexical structure of Tone 3 words with a 'Tone 3' label.161
Figure 5.1. Constraint ranking for coda neutralization and assibilatory affrication in Korean..171
Figure 5.2. Summary of surface variations of suffixed forms from different stem-inputs...187
Figure 5.3. Grammar changes with stem-final variations in Korean......................187
Figure 5.4. Target constraint ranking for standard pronunciations in Korean......193
Figure 5.5. Stem-final [s] proportions over ten learning generations..197
Figure 5.6. Constraint values of five constraints and the corresponding constraint ranking at the end of the M stage simulation in Generation 1..........................198
Figure 5.7. Constraint values of five constraints and the corresponding constraint ranking at the end of the M stage simulation in Generation 5..........................199
Figure 5.8. Development of constraint values of the selected five constraints.......199

Figure 5.9. Constraint values of four constraints and the corresponding constraint ranking at the end of the M stage simulation in Generation 1..........................200
Figure 5.10. Constraint values of four constraints and the corresponding constraint ranking at the end of the M stage in Generation 5...201
Figure 5.11. MEP[X+Ø]$_{\text{BARE}}$ and the mean of other MEPs over ten learning generations...202
Figure 5.12. SP development of surface allomorphs of 'sickle' over ten learning generations...204
Figure 5.13. SP development of surface allomorphs of 'day' over ten learning generations...205
Figure 5.14. SP development of surface allomorphs of 'flower' over ten learning generations...207
Figure 5.15. SP development of surface allomorphs of 'clothes' over ten learning generations...209
Figure 5.16. SP development of surface allomorphs of 'breast' over ten learning generations...211
Figure 5.17. SP development of surface allomorphs of 'face' over ten learning generations...212
Figure 5.18. SP development of surface allomorphs of 'field' over ten learning generations...213
Figure 5.19. SP development of surface allomorphs of 'red bean' over ten learning generations...214
Figure 5.20. SP development of surface allomorphs of 'outside' over ten learning generations...217
Figure 6.1. Lexical association between six [bɛd-ən] tokens.................................227
Figure 6.2. Lexical associations between [bɛt] and [bɛd-ən] tokens in the P stage (solid lines) and the M stage (dotted lines)...228

Figure 6.3. Output error rates of [bɛd-ən] and [vut-ən] from the 100th learning cycle to the 10,100th learning cycle in the M stage simulation....................................241

List of Acronyms

ATF	Adjusted Token Frequency
AP	Alternating plural form
CDS	Child-directed speech
ERC	Elementary Ranking Condition
GLA	Gradual Learning Algorithm
HG	Harmonic Grammar
HS	Half Sandhi
IEP	Individual Error Proportion
IO	Input-Output (Correspondence)
LIOO	Lexically-indexed Output-Output (Correspondence)
MEP	Morphological Error Proportion
NAP	Non-alternating plural form
NPL	Naïve Parameter Learning
OCP	Obligatory Contour Principle
OO	Output-Output (Correspondence)
OT	Optimality Theory
PSI	Probabilistic Selection of Input
SCC	Strict Cycle Condition
SP	Selection Probability
SURA	Single Underlying Representation Assumption
T3S	Tone 3 Sandhi
UR	Underlying representation

Chapter 1

Introduction

An abstract and economic lexicon has long been recognized as the core of morphophonology from the view of generative linguistics (e.g. Chomsky & Halle 1968; Kenstowicz & Kisseberth 1977, 1979; Kiparsky 1973, 1982a), and one primary reason for the assumption is that we prefer economy ceterus paribus, which has been conceptualized as Occam's razor. Thus, if we compare a lexical economy model and a rich lexicon model which make exactly the same predictions, the former seems more preferable with the above rule of thumb. Nevertheless, the reasons to cast doubt on lexical economy are twofold. The first arises from the neurobiological facts: As questioned in Krämer (2012:4), since we have an estimated 100 billion nerve cells after all, "why should our mind be so obsessed with economy when it comes to storage?" Secondly, Occam's razor may not be applicable to the comparison between the two assumptions of the lexicon since they may never make exactly the same predictions. For example, lexical abstractness predicts a lower consumption of the lexical storage but lexical richness on the other hand reduces the burden on the computational capacity (by avoiding computing abstract lexical representations), and the two types of 'economy' are not directly comparable in Occam's razor. If we take these issues into consideration, the premise of lexical economy, which forms the basis of the majority of generative models, seems rather questionable. This thesis, by contrast, follows psychological evidence suggesting a rich lexicon to build blocks of a morphophonological learning model and seeks to explain morphophonological acquisition with transparent lexical processes.

In this model, we assume that learners exploit the privilege of having a potentially rich memory space, as found in recent research on (morpho-)phonology, and store all surface allomorphs of a morpheme faithfully with fine phonetic details in their lexicon as possible input options of the morpheme. Different output patterns at sequential stages in morphophonological acquisition are accounted for with gradual probability changes in selecting different allomorphs as the input in this model (thus Probabilistic Selection of Input; PSI), and diachronic changes occurs as learners of each generation prefer different allomorphs as inputs. In a nutshell, we set out to construct a morphophonological acquisition model that has a stronger explanatory power driven by its neurobiological foundation, and we hope to explicitly spell out the 'human mind' process that lies behind morphophonological acquisition.

This chapter serves as a preface that draws the big picture of the above claim, starting with a review of arguments in favor of lexical abstractness in §1 and challenges to this lexical abstractness hypothesis from §2 to §4. In §5, the false impression that a rich lexicon and a phonological grammar are mutually exclusive will be dismissed. Recent applications of an input selection process will be reviewed in §6, and a 'hybrid' model with a constraint grammar is introduced in §7. Finally, the guidelines that will be followed in the design of PSI are specified in §8.

1. Lexical abstractness

Various phonological models have been proposed since Chomsky & Halle's *The Sound Pattern of English* in 1968, which necessarily include two components – lexicon and grammar. The majority of generative phonology has focused on the importance of the phonological grammar, dealing with (morpho-)phonological alternations and phonotactics by a set of SPE-style rules or constraints in constraint-based model. The importance of the lexicon, on the contrary, has attracted less attention from generative phonologists. In a rule-based system, the interaction between lexicon and grammar was largely addressed in Lexical Phonology (Kiparsky 1982b, 1985) with a multi-leveled phonological grammar encoded with different lexical and post-lexical rules at separate levels. In constraint-based models, the interaction is represented with a group of constraints that specifically refer to lexical information (e.g. lexically-indexed constraint; Pater 2000, 2009) or lexically-conditioned constraint rankings (e.g. Antilla 1997, 2002). The lexicon, in these models, serves as a reference for the construction of lexically-specified rules or constraints or a variable which can generate multiple grammars based on different lexical categories. Otherwise, the lexicon has been assumed as simple-structured and highly-economic repository that stores no redundant information at all.

Halle (1985:105) strongly endorses the idea of an abstract lexicon by arguing that "one may speculate that space in our memory is at a premium", whose reasoning comes from regular English stress patterns and onset cluster phonotactics. Halle claims that English speakers by no means store stress information along with the words in the lexicon since the primary stress mostly falls on the antepenultimate syllable. The speakers therefore can induce a phonological rule from the major stress patterns with some exceptions (e.g. statuétte, devélop, órchestra) and eventually project abstract representations without stress. Russian loanwords in English play an important role in Halle's argument in favor of an abstract lexicon as well; in Russian loanwords, stress follows the antepenultimate stress rule in English, which may result in a stress shift from the original stress position in the source (e.g. bolshevík in Russian but bólshevik when borrowed to English). Halle questions, if stress patterns are simply memorized as they are perceived by English speakers, the stress shift should have never occurred in the Russian loanwords. The judgments about legal onset clusters in English also inspired Halle to raise the questions to a rich storage of surface forms. The number of legal onset clusters is around thirty in English, and Halle suggests the assumption that English speakers just memorize these clusters and exhaustively check them every time to be impractical. On the contrary, Halle believes that English speakers do have 'very clear intuitions' about whether words like 'bnin' or 'vlim' are English-like (for evidence from recent experimental studies, see Hayes & Wilson 2008 and Hayes & White, 2013). By 'very clear intuition', Halle refers to some general principles acquired by English speakers. In sum, these facts prompt Halle to suggest that regular patterns are induced as phonological rules, and underlying representations (UR) are abstract away from these predictable patterns.

Among the generative models with an abstract lexicon, some hypotheses, such as underspecification (e.g. Archangelli 1988; see Colina 2013, Lahiri & Reetz 2002, 2010, and Scharinger & Idsardi 2014 for recent supportive arguments), allow the projection of highly abstract URs, leaving the assignment of phonological features entirely to a grammar with contextual alternation rules and default elsewhere rules (e.g. assimilation rules for the feature agreement of [Lab] and [Dor] and a default rule assigning the unmarked feature [Cor]).

To an extreme, some generative models were developed with a lexicon that can arbitrarily project and store non-surface-true phonological elements in URs; grammar is once again responsible for creating or learning appropriate rules to derive target outputs. [1] Hyman (1970) characterizes such non-surface true phonological elements as 'absolute neutralization'. The alternation of stem-final /t/ in Tsuut'ina (also Sarcee) is considered one example, in which the stem-final /t/ does

[1] E.g. Archiphoneme (e.g. Jakobson 1929 and Trubetzkoy 1929, 1936), which can be posited for neutralization without alternation, allowing more arbitrary lexical abstraction. Also see the Nupe and Tsuut'ina examples below.

not surface faithfully in any context; the segment surfaces as [t’] before a vowel-initial suffix and as [ɬ] before [a], and is deleted elsewhere (Sapir 1933; quoted in Kentowicz 1994). In other words, although /t/ never occurs in any surface form, the abstract phoneme must be posited in the lexicon from which other surface coronal variants are derived. Another example of absolute neutralization comes from the Nupi derivations in (1): Assimilation rule labializes and velarizes the intervocalic consonant when the second vowel is [-low] as in (1a) and (1b), whereas a [+low] vowel in (1c) blocks the application of the assimilation rule. However, labialization still occurs in (1d) although the second vowel is not [u] as in (1a). The difference between (1c) and (1d) leads Hyman to posit a non-surface-true /ɔ/ in the UR of (1c) (i.e. /egɔ/), and the assimilation rule applies first and is then counterbled by a vowel lowering rule (i.e. /egɔ/→//egʷɔ//→[egʷa]).

(1) Assimilation process in Nupe

a.	/egu/	→	[egʷu]	‘mud’
b.	/egi/	→	[egʸi]	‘child’
c.	/ega/	→	[ega]	‘stranger’
d.	/ega/?	→	[egʷa]	‘hand’

With the above and many other arguments in favor of an abstract and economic lexicon, for decades, morphophonology has been studied in accord with this assumption of a minimum consumption of the mental space. Surface allomorphs of a morpheme in different morphological contexts are thus derived from a single UR without predictable elements (henceforth the Single UR Assumption, or SURA) in such generative frameworks. The abstract level of a morpheme’s UR depends on the model adapted by a researcher, which can be as transparent as being one of its surface allomorphs, or can contain underspecified or non-surface true segments as reviewed above. While these models may be analytically adequate as they have no difficulty in demonstrating how the lexicon feeds abstract lexical items to a phonological grammar and surfaces as the intended output forms, learnability of these models has rarely been put to any empirical test to demonstrate the transition from some initial, perhaps more detailed (see §4.1 below), lexical representations in child (morph)phonology to more abstract representations in adult (morpho)phonology. Hyman, for instance, did express his belief in the learnability of absolute neutralization, but only vaguely attributed the learnability to children's capability of 'reasoning':

> "…it is claimed that a child need not hear the phonetic shape of an underlying segment to have stored it in his brain…the child does not learn an abstract underlying representation solely from the phonetics of the individual morphemes…but rather has the additional ability to 'reason', instead of merely mimicking." (1970:76)

The advent of constraint-based theories allows the learnability of (morpho-)phonology to be easily tested in computational modeling, including SURA. Some previous machine learners demonstrated that abstract URs were learnable (e.g. Jarosz 2006a, Merchant & Tesar 2005, Tesar 2008, Tesar & Prince 2003, Tesar et. al 2003), but empirical predictions of morphophonological acquisition were still absent in these approaches as the training inputs were limited to toy data.

The difficulty for development an efficient learning algorithm of URs suggests that lexical parsimony in fact sacrifices computational economy, and that a language model with an economical lexicon is by no means simpler as a whole. Yip (1996) shares the same thought and suggests that learning abstract URs might be a difficult acquisition task, and "an easier strategy may be to stick close to the phonetic surface form". When possible inputs are restricted to surface-true allomorphs, the learning model has a finite search space as learners only have limited choices in the input selection process. If the learners are allowed to posit highly abstract representations, the computation space is in theory infinite. Learning strategies might be invented to restrict lexical abstractness,[2] but it can only at best eliminate a small proportion of arbitrary abstract representations; the remainder of the search space still costs great computational efficiency for very limited goals. Furthermore, as we have a great number of brain nerve cells to process and store information, it is less convincing that speakers do not exploit this privilege when learning morphophonology (or a language system in general) or even after their lexical and grammar generalization become stabilized.

Summing up, given insufficient empirical evidence that the lexicon must be abstract and limited, we should hesitate to take SURA for granted when building a morphophonological theory. The remainder of this chapter further offers a closer

---

[2] For example, Inkelas (1994) proposes that underspecification is limited to segments participating alternations.

examination of SURA, including a few challenging learning problems that demonstrate the feasibility and necessity of abandoning the assumption for a morphophonological acquisition model with a rich lexicon.

2. Circularity Problem

I begin my examination of models with an abstract lexicon by indicating an intrinsic difficulty in acquiring morphophonology with SURA, which lies in the mutual dependency between learning the lexicon and learning the phonological grammar. A simple example can be illustrated with the hypothetical patterns in (2). If the first two learning inputs are [bap] and [ba.ba], learners cannot decide whether the alternation surfaces from final-devoicing or intervocalic voicing. A markedness-based criterion may not be a helpful since one phonological process might be more marked on one dimension but less marked on another.[3] The choice between the two rules thus depends on which surface allomorph (/bap/ or /bab/) is acquired as the stem UR. Without knowing which rule is the target grammar, however, learners can never deduce the stem UR from the derivation /bab/→[bap]. This mutual dependency issue is dubbed as Circularity Problem in Albright & Hayes (2011):

> "The optimal choice for the URs depends on having a reasonably good hypothesis about the grammar, but the grammar cannot be formulated without a hypothesis about the set of input → output mappings that it must perform." (2011:674)

(2) Hypothetical final devoicing alternation

[bap]	'Stem A (sg.)'	[ba.ba]	'Stem A (pl.)'
[bat]	'Stem B (sg.)'	[ba.ta]	'Stem B (pl.)'

The pair [bat]~[ba.ta] appears to be counterevidence against the intervocalic voicing rule, but a few of these tokens only help reject an obligatory rule, not the possibility of an optional rule. For the final devoicing rule to win the competition between grammar assumptions, learners must observe a gradually growing number

[3] For example, an intervocalic voiceless obstruent may be phonetically more marked, but a voiced obstruent coda may be typologically more marked. See Haspelmath (2006) for arguments against universal phonological markedness.

of exceptions to the intervocalic voicing rule. Before the learners reject grammar assumptions with a great number of exceptions, the two UR assumptions should also be stored in the lexicon and retrieved by learners who switch back and forth between correspondent grammar assumptions in a gradual learning process. It is of course possible to assume that multiple allomorphs are stored before URs are acquired and that all but one are abandoned once learners converge on some UR assumption. I propose that the memory of language, akin to other episodic memories, gradually fades away (see §4.2), but the speed of memory decay, either steady or unstable due to the interference by other activities (see Hardt et al. 2013 and references cited therein), cannot be controlled by language learners. The primary implication here is, language learners do not consciously abandon lexical information either during or after the main course of language acquisition.

3. Relexicalization and rule inversion

We now temporarily set the learning problem aside, assuming that there is a model which can always acquire a target phonological grammar along with correct URs simultaneously. This model, however, may in fact be too powerful to be the ultimate solution. We expect a model that analyzes same morphophonemic alternations to conclude with the same 'correct' generalization consistently, but a real morphophonological learning process may be less 'perfect'; learners may adopt different lexical representations for a morpheme over time, which results in the emergence of an 'inversed' phonological grammar. Vennemann (1972) dubs the former as 'relexicalization' and the ensuing emergence of the inversed grammar as 'rule inversion'.

An example of total rule inversion is presented in (3): In Stage 1, a form *B* is derived from *A* in the environment *D*, which is a morphosyntactic domain such as singular, plural, etc. If the environment *D* is accidentally considered a more basic domain by learners of the next generation, *B* will be identified as a more basic form. Stage 2 then occurs to derive *A* from *B* in the environment $\underline{D}$, which is a complementary morphosyntactic domain.

(3) Total rule inversion

Stage 1: A→B/D

Stage 2: B→A/$\underline{D}$

An example of total rule inversion provided in Vennemann (1972) is the diachronic change of the indefinite article 'an/a' in English. Since the article was developed from the clitic use of 'one', it should be straightforward to associate *an* as the reduced form of 'one', and the allomorph 'a' is derived from 'an' with Stage 1 in (4). Evidence for moving to Stage 2 lies in the acquisition data in which children use 'a' earlier (e.g. 'a' apple) and the description in contemporary dictionaries depicting 'a' as the citation form. See also Churma (1982), Hayes (2000), Martinez-Gil (1997), and Vennemann (1974) for cross-linguistic rule inversion patterns.

(4) Nasal dropping rule (Vennemann 1972:213)

Stage 1: $/an/_{Art} \rightarrow [a] / \_\_\_\_ \#C$

Stage 2: $/a/_{Art} \rightarrow [an] / \_\_\_\_ \#V$

Total rule inversion may also be extended to the final devoicing data (2) as follows. In Stage 1, learners can acquire /bab/ as the more basic form of Stem-A and the target rule $/b/ \rightarrow [p] / \_\_]\sigma_{singular}$, provided that morphosyntactic context is always part of the rule. For reasons that will be elucidated later, learners of a new generation consider singular the more basic domain and the allomorph /bap/ as the more basic allomorph of Stem-A. Stage 2 of rule inversion $/p/ \rightarrow [b] / \_\_]\sigma_{plural}$ thus occurs to replace a final devoicing rule with an intervocalic voicing rule.

Rule inversion raises the problem for SURA models since learners must first retain multiple allomorphs into their adulthood to produce surface variation. The diachronic change is then initiated by passing the variation patterns to learners of the next generation as learning inputs. Moreover, one criterion used to determine the basic allomorph of a morpheme is the token frequency of surface allomorphs (see §4.2 below), which requires a rich storage of surface allomorphs to keep tracking the token frequency of each allomorph on-line to decide the basic allomorph at any given point. All in all, rule inversion and other learning issue reviewed in previous sections seem to generally rule out SURA as a plausible morphophonological model. This local conclusion leads to further discussions of more evidence for a rich lexicon below in §4.

4. Evidence for a rich lexicon

Kirchner (1999) and Port (2007, 2010) suggest some extra-linguistic factors that could further motivate the assumption of an abstract lexicon. One might be the severely limited computer memory capacity in the past, providing minimal support for the storage of surface forms. Another cause is perhaps the requirement of doing linguistic analyses and constructing formal language theories with discrete symbols or logistic formal expressions, i.e. a system that describes (abstract) categories and thus excludes most, if not all, gradient phonetic details or non-categorical lexical information. Following this tradition in the past were the frequent uses of abstract phonological elements such as phoneme, syllable, and economic lexicon in psychological and psycholinguistic research as well (e.g. Dell 1986; Cutler & Norris 1988; Fear et al. 1995).

Direct evidence, such as the performance in various experiment tasks and a great number of natural language patterns described below, yet strongly suggests that the lexicon is considerably richer than the one assumed in traditional generative theories. For example, previous psychological research suggests that general and specific episodic knowledge is stored in the same rich memory space (e.g. Anderson & Ross 1980; Graf & Schacter 1985; see Tenpenny 1995 for a comprehensive review). In light of this, it is highly possible that abundant, and perhaps even redundant, language information is encoded in a rich lexicon. More specifically, by 'rich lexicon' in this thesis, I propose a lexicon that stores individual paradigms (including surface allomorphs), the token frequency of a paradigm, and phonetic details of a paradigm. The goal of §4 is to summarize findings in favor of exploiting the lexical space, which particularly contradict two major hypothesis in SURA, i.e. lexical abstractness and single UR.

4.1 Against abstractness: Phonetic details in child and adult phonology

This section focuses on psycholinguistic evidence that seems to strongly suggest the storage of individual paradigms with rich phonetic details, which serves as a counterexample of lexical abstractness allowed in SURA.

As discussed in §1, some SURA models allows non-surface-true segments to appear in the single UR of a morpheme, which implies that learners can always reason abstract URs. This assumption can be falsified if it is shown that learners in fact encode as many surface phonetic details as possible in their lexical representation as suggested in Yip (1996), which was likely found in previous research on child and adult phonology. In child phonology, for example, babies can associate their own vocalization with their own articulatory gestures, creating an auditory-articulatory 'feedback loop' (e.g. Fry 1966, Stoel-Gammon 1998), and those frequently vocalized sequences can assist them to identify similar phonetic forms in the learning inputs, which usually become their first few words later. If the vocalization outputs and their phonetic details are not stored in the lexicon for comparison, the self-training, reinforcement, and bootstrapping processes seem very unlikely (see also Stoel-Gammon 2011 for a detailed review). Swingley (2003) and Swingley & Aslin (2002) also conducted experiments focusing on 17-month-old children's eye fixation when presenting correct pronunciations and minimal mispronunciations. The results in Swingley (2003) suggest that the subjects looked at the target pictures longer when corresponding correct pronunciations were presented and that young children encode sufficient rich phonetic details for distinguishing phonetic forms. In particular, the correct pronunciations were reported to have no known neighbors in the subjects' lexicon, thus excluding the interpretation that the phonetic details of correct pronunciations can only be encoded following the expanded lexicon with more neighbors (e.g. Charles-Luce & Luce 1990; Fowler 1990). More solid evidence may be grounded on infants' capability of identifying different talkers from learning inputs. Juscyzk et al. (1992) found that 2-month-old infants can distinguish same syllables produced by different speakers which they perceive during a habituation phase (see also Kuhl 1979, 1983). This result can have a straightforward account by assuming that tokens produced by different speakers are stored faithfully in the infants' lexicon.

Not surprisingly, perhaps, there have been studies against the above findings. For example, Ferguson & Farwell (1975), Walley (1993), Waterson (1971), and many others, following the observation that young children use the same output for multiple words and fail to identify minimal mispronunciations, also claim that phonological representations at early stages might be holistic and encoded with merely a few phonological features rather than full segmentation. Swingley (2003) and Swingley & Aslin (2002) nevertheless argue that children's phonological representations cannot be determined solely upon their production outputs or performance of linguistic judgments; production outputs and their lexical representation may greatly mismatch, and linguistic judgments might be influenced by meta-linguistic skills. There could also be alternative interpretations of the experimental results: Infants may simply store phonetic details in some auditory memory and may have never transfer the information into the grammatical lexicon; these phonetic details therefore are never relevant in (morpho)phonological acquisition. This assumption naturally begs the question why learners cannot store phonetic details into the grammar lexicon later as well. Moreover, research into adult phonology indicates that phonetic details are indeed stored and used grammatically in linguistic tasks such as lexical decision.

Goldinger (1996) tested whether adult English speakers could encode different sounds in the memory and then recall them from the memory (see also Craik & Kirsner 1974; Goldinger et al. 1991). 360 subjects were first grouped as three participation pools to listen to 150 words produced by two, six or ten male or female speakers. They were not informed another delayed identification test that would require them to recall the words they had heard so they had no reason to focus on memorizing the sounds and their talker-specific phonetic details they perceived. Thus, if they did recall these words more easily upon hearing the tokens produced by the same speakers in the second task, they might have encoded the phonetic details by default to facilitate the retrieval of talker-specific lexical entries.

After different delay lengths (five minutes, one day, and one week), the second task was implemented, in which the same subjects were asked to listen to 300 words, which were either in a voice they had previously heard or in a new voice. The subjects' task was to identify whether each word they heard in this session is an old word they had heard in the first session, or a new word that only appeared in the current session. The results suggest that when a word was in an old voice, the subjects were more likely to recall the word as an old word. This tendency was not affected by the number of speakers producing those words even if the 150 words were produced by ten different speakers, which also suggests a rich lexicon which does not only store very limited voice information.[4]

The experimental results of learning artificial morphophonemic alternations in Peperkamp & Dupoux's (2007) might be potential evidence against Goldinger's findings. In their experiments, Peperkamp & Dupoux trained adult subjects to learn an artificial language with a phonemic stop voicing contrast and an allophonic fricative voicing contrast or a phonemic fricative voicing contrast and an allophonic

[4] The evidence that memory plays an important role is also revealed by the inversed relation between the recall rate of the words presented in their original voice and the delay length; the sounds stored in the memory fade away after they have not been activated for days. This memory decay effect will also be introduced into the proposed learning model in §1.4 of Ch. 2.

stop voicing contrast. The prediction is that if the morphophonemic alternation is acquired, the subjects should not be sensitive to an allophonic voicing change since allophones correspond to one underlying (abstract) phoneme. Conversely, if phonetic details are always encoded in lexical representations, phonemic and allophonic voicing changes should be equally noticeable. The first prediction seems to be borne out with the experimental results demonstrating unperceivable allophonic voicing changes. However, whether phonetic changes are perceivable is not necessarily the consequence of storing phonetic details or not. It could still be the case that the phonetic details of allophonic contrasts are stored but mapped to the same category; allophonic variants are thus perceived as the same phoneme.

Summarizing, with evidence strongly pointing to a rich lexicon with fine phonetic details, it should be safe to conclude that learners prefer encoding lexical representations highly similar with or identical to surface representations, rather than 'reasoning' abstract lexical representations. SURA models that permit extremely abstract URs as in absolute neutralization or underspecification, are thus considered implausible.

4.2 Against parsimony I: Lexical entries and token frequency

The empirical evidence against lexical abstractness does not reject the hypothesis that there could still be only one single UR for every morpheme. The successful identification of talker-specific tokens as reviewed previously nevertheless allows us to ask, if speakers do not hesitate sparing their lexical memory to store relatively redundant lexical entries, why do they have to induce single URs rather than simply store surface allomorphs? This section introduces performance evidence favoring allomorph listing and thus against lexical parsimony.

In an approach assuming a rich lexicon that stores individual surface forms separately, it is possible for each of these forms to have its own token frequency,[5] which allows word-specific production and recognition patterns. One well-known token frequency effect is that high-frequency lexical items may be subject to lenition more easily as observed in two oft-quoted examples of word-specific segment deletion in English: First, a schwa in high-frequency words such as 'every' is deleted

[5] In an Exemplar-based model that stores every phonetic variant as in Kirchner et al. (2010) and Johnson (1997), the token frequency of a word can be tracked by counting the number of the word's phonetic variants. The two ways of counting token frequencies bear no difference in the core of this thesis (but see Ch. 6).

more frequently than in low-frequency words such as 'mammary' (Hooper 1976 and Bybee 2000). Second, a word-final /-t/ is largely omitted in high-frequency words such as 'just' and preserved in low-frequency words such as 'jest' (e.g. Bybee 2000, 2002, Coetzee & Kawahara 2013). In addition, Gahl (2008) also discovered a significant duration difference between high-frequency and low-frequency homophones, and the latter are usually considerably longer.

Word-specific token frequency effects have also been observed among morphologically related words if we assume that these paradigms (and thus allomorphs of a morpheme) are listed individually in the lexicon with their own frequency (cf. SURA). Following the general fact that high-frequency lexical items are retrieved faster, the rich memory model makes the prediction that the token frequency of whole morphologically complex forms, rather than their base, determines how fast they can be accessed. The prediction is borne out in the lexical decision studies of English past tense *–ed* forms (Hare et al. 2001), Dutch plural forms (Baayen et al. 2003), and German nominalized *–ung* forms (Clashen & Neubauer 2010).

Allomorph listing with individual token frequency in a rich lexicon is also consistent with the claim that token frequency serves as one major factor in determining basic allomorphs in the development of paradigmatic leveling. In the development of paradigmatic leveling, more frequent paradigms are extended to replace less frequent paradigms, but not vice versa (Mańczak1980:284). Furthermore, as mentioned in §2 and §3, all of the allomorphs must be stored in the lexicon, so basic allomorphs can be determined based on the frequency information at any given point; learners continue to receive learning inputs and update their frequency information, and basic allomorphs may change after each update. For example, assuming that the selection of a basic allomorph relies only on token frequency, if at one stage allomorph *A* accidentally appears 54 times in the learning data, and *B* only 46 times, *A* is identified as the basic allomorph. At the next stage the distribution may be 86 times for *A* and 114 times for *B*, at which point *B* is considered the basic allomorph. Put differently, there is no one-time decision on what basic allomorphs are, and surface allomorphs should be retained in the lexicon to allow the flexibility of switching basic allomorphs. If basic allomorphs not only change temporarily within a learning generation but also shift between learning generations,[6] relexicalization occurs (§3).

[6] Patterns that can be observed in child phonology often occur in diachronic sound changes. Other

Korean is an example of how relexicalization is triggered by selecting high-frequency allomorphs as basic allomorphs. Korean is known to have abundant coronal obstruent contrasts /t, tʰ, t’, tʃ, tʃʰ, tʃ’, s, s’/. In the coda position, however, they are neutralized as an unreleased stop [t̚] as in (5) below. A standard generative analysis would simply conclude that the accusative allomorph is the UR of each stem morpheme, whose stem-final obstruent is neutralized in a non-suffixal context.

(5) Korean coda coronal neutralization (Martin 1992)

a.	[pat̚]	‘field’	[patʰɨl]	‘field (acc.)’
b.	[tʃʌt̚]	‘milk’	[tʃʌdʒ-ɨl]	‘milk (acc.)’[7]
c.	[k’ot̚]	‘flower’	[k’otʃʰ-ɨl]	‘flower (acc.)’
d.	[ot̚]	‘clothing’	[os-ɨl]	‘clothing (acc.)’

However, surface variations of accusative forms have been discovered recently (e.g. Martin 1992, Kang 2003, Jun & Lee 2007, 2010) as demonstrated in (6). The variation patterns are not random; they include the available alternations from other morphologically related pairs.

(6) Extended alternation in Korean (Martin 1992)[8]

a.	[pat̚]	‘field’	[patʰɨl], [patʃʰɨl], [pasɨl]	‘field (acc.)’
b.	[tʃʌt̚]	‘milk’	[tʃʌsɨl], [tʃʌdʒɨl]	‘milk (acc.)’
c.	[k’ot̚]	‘flower’	[k’otʃʰɨl], [k’osɨl]	‘flower (acc.)’
d.	[ot̚]	‘clothing’	[osɨl]	‘clothing (acc.)’

examples unrelated to morphophonology include variations caused by misperception (e.g. Ohala 1975, 1978; Ohala & Lorentz 1977), which can be found in both early phonological and long-term diachronic developments. In Achinese, Toura, and Mbay, a post-vocalic velar nasal /ŋ/ is developed from a nasalized vowel (i.e. Ṽ→Vŋ) because perceptually the two sound sequences are similar to each other. The same patterns were found in child language data of French, Japanese, Mandarin, and Yucatec Maya as the Ṽ and Vŋ sequences also frequently substitute each other (for more details, see Greenlee & Ohala1980 and the references cited therein).

[7] Coronal plain (unaspirated) stops and affricates are voiced in an intervocalic position in Korean (e.g. Cho 1967, Yun & Jackson 2004).

[8] The surface variation of ‘clothing (acc.)’ was not documented by Martin (1992), but there was one token of [otʃʰɨl] in Jun & Lee’s (2007) elicitation experiment (Exp I). The minor difference can be either incidental or a continuous diachronic change from 1992 to 2007.

The [t]~[s] alternation in (5d) is extended as shown in [pas-ɨl], [tʃʌs-ɨl] and [k'os-ɨl] in (6a), (6b), and (6c), and the form 'field (acc.)' also has the variant [patʃʰ-ɨl] in (6a) with the extended [t]~[tʃʰ] alternation. One plausible view is that the bare form of each stem morpheme is considered the basic stem allomorph, and a phonological grammar with a set of probabilistic rules (7) is acquired by Korean speakers.

(7) Probablistic rules acquired with bare allomorphs as basic forms (cf. Jun 2007)

a. /t/→[s]/___+ɨl

b. /t/→[tʃʰ]/___+ɨl

Evidence supporting this view is Lee's (1999) research on the Korean morphosyntactic acquisition, which reveals the fact that in child-directed speech, bare forms occupy 75% of the learning inputs, 20% of the inputs are nominative forms with the high front vowel suffix [-i], and only 5% have other types of suffixes. Even if we assume that the 5% are accusative forms, a high frequency ratio (15:1) still represents a distribution extremely biased toward bare form allomorphs. Ch. 5 aims at modeling this diachronic development in Korean with the same selection process that determines the basic allomorph of a morpheme.

The above evidence suggests that individual surface forms, including morphologically related paradigms, are stored separately as individual lexical entries. In this view, all allomorphs, rather than simply one single UR, of a morpheme are listed in the lexicon with their token frequency, which is clearly against lexical parsimony as assumed in SURA.

4.3 Against parsimony II: Multidirectional paradigmatic leveling

My final review of empirical evidence supporting the storage of surface allomorphs concentrates on morphophonological outputs constrained by the shape of multiple non-base allomorphs (i.e. multi-directional paradigmatic leveling).

Previously paradigmatic leveling was assumed to be unidirectional; all the derived forms like suffixed forms are leveled to their base paradigms such as bare stems (e.g. 'damn' [dæ̃m]~'damning' [dæmĩŋ], 'damn' *[dæ̃mn]~'damning' [dæmnĩŋ]; Benua 1995, 1997). It is thus possible to assume that only base paradigms are stored in the lexicon, which all morphologically related forms are faithful to. This approach nevertheless runs into problems in dealing with inflectional morphology not only because surface paradigms are very unlikely to be derived from one single base paradigm but also because one surface paradigm might be leveled to multiple paradigms as in Classic Arabic, which is not possible if non-base allomorphs are excluded from the lexicon.

Classic Arabic has a template restriction that verb stems must end with CVC in morphologically complex forms; noun stems can instead end in CVC, CV:C, and CVCC when following by a suffix. By contrast, the left edge of noun stems can never begin with CCV, whereas verb stems are free to have such an initial cluster. McCarthy (2005) proposes that the template of verb stems and the free occurrence of a stem-initial CCV cluster emerge from the paradigmatic leveling between multiple paradigms with various inflectional affixes.

First of all, all inflectional noun suffixes are vowel-initial, but inflectional verb suffixes can have either an initial vowel or consonant (see McCarthy 2005:179-180 for a complete list). With a consonant-initial suffix CV, a stem-final CV:C is impossible since the form such as *CVC.<u>CV:C</u>-.CV contains a word-medial superheavy syllable, which is illegal in Classic Arabic. Put differently, all verb forms with a CV suffix have a stem with a stem-final CVC, and all morphologically related forms are leveled to these verb paradigms to create the right-edge verb stem template. By contrast, since noun suffixes are always vowel-initial, a stem-final cluster CV:C or CVCC is possible in suffixed forms such as CVC.CV:.C-V or CVC.CVC.C-V. Without a consistent stem-final cluster, even if all noun paradigms are leveled to suffixed paradigms, no single noun template may emerge.

Verb paradigms are also leveled to their prefixed forms to allow a stem-initial CCV cluster. Since verbs in Classic Arabic can have a CV prefix, the first consonant in a stem-initial cluster CCV is always syllabified as the coda of the word-initial syllable as in the forms like CV-C.CV…. To maintain paradigmatic uniformity, a stem-initial CCV cluster is allowed in non-prefixed verb paradigms. Nouns in Classic Arabic nevertheless do not have the same prefix to allow a stem-initial CCV cluster in any noun paradigm; the cluster cannot be introduced into morphologically related noun paradigms via paradigmatic leveling.

The above analysis suggest that both prefixed and suffixed verb paradigms must be listed in the lexicon as the reference of paradigmatic leveling, which challenges lexical parsimony in SURA directly. Another example of multidirectional paradigmatic leveling is French liaison (Steriade 1999, 2001a). In French liaison, the phrase 'next-Masc. stop' can be produced as [pʁɔʃɛ̃n aʁe], in which the nasalization in the stem is borrowed from the masculine paradigm [pʁɔʃɛ̃] and the stem-final nasal from the feminine paradigm [pʁɔʃɛn]. Positing a single UR such as /pʁɔʃɛn/ for the three paradigms is not feasible since the derivation from the UR to the masculine paradigm requires a nasalization rule (i.e. /pʁɔʃɛn/→//pʁɔʃɛ̃n//) and a nasal deletion rule (i.e. //pʁɔʃɛ̃n//→[pʁɔʃɛ̃]). However, as both rules do not apply in any other phrasal contexts in French (Burzio 2005:70), this analysis does not provide any thorough explanation for the phonological system of French; after all, one can in theory invent individual grammars to derive all surface forms from any arbitrarily assumed URs. Accordingly, Burzio (2002, 2005) also argues in favor of a rich lexicon that stores a bare form and its morphologically complex forms which altogether influence the production of an output.

Multidirectional paradigmatic leveling can also account for the vulnerability to phonological alternations in different morphological contexts. In English, Level 2 suffixes can shift the initial stress on the root to the penultimate syllable, but the novel form 'remédiable' (cf. 'rémedy') is possible and *'paródiable' (cf. 'párody') is not. The difference lies in the paradigms of the two stems; the former has two paradigms with different stress locations 'rémedy' and 'remédial', but the latter has only one 'párody'. To account for this difference, Steriade (2001a) proposes that the novel form must be faithful the lexical paradigms sharing the same stem, which is satisfied by 'remédiable' since the novel form and the existing form 'remédial' have the same stress location. To the contrary, *'paródiable' is considered impossible due to a stress shift from the only lexical paradigm 'párody'. Put differently, the novel form 'remédiable' is possible because it applies Level 2 stress shift without losing paradigmatic uniformity between the output and any of the stored related paradigms (and stem allomorphs). The difference between *'paródiable' and 'remédiable' thus cannot be explained without a rich lexicon with all surface paradigms.

5. Rich lexicon and phonological generalization

The proposal of a rich lexicon naturally raises the question why learners do not simply memorize every learning input and reproduce them faithfully without any level of an abstract grammar generalized from learning inputs, as argued in Halle (1985). However, it is a false belief that a rich lexicon makes grammar redundant. For example, we can encode and thus identify various details of the appearance of eagle, parrot, and albatross but can still generalize the common features such as two wings, two legs, and a beak. When we need to categorize a previously unidentified creature with all these features, we will probably assign it to some bird categories. Pierrehumbert (2006) and Guy (2007, 2014) suggest that a pure lexical generalization model (e.g. Exemplar Theory) is not completely tenable since it cannot account for the extraction of abstract rules or constraints as shown in child and adult phonology. For example, Berent (2013) and Berent, Marcus, Shimron, & Gafos (2012) claim that phonology must be an algebraic system projecting 'relations' from observable inputs and apply these relations to novel words without any lexical precedents (i.e. no lexical generalization of novel words can be drawn). Evidence verifying this claim is the restriction in the Hebrew word-formation, which prohibits XXY but not XYY consonant sequences (e.g. *sisem vs. simem), which can be extend to novel words including non-Hebrew consonants such as [θ] (Berent, Marcus, Shimron, & Gafos 2012). That is, Hebrew speakers do not simply acquire discrete restrictions that forbid first two identical native consonants, but rather project a more general code composed of algebraic variables which are applied across the board. In sum, a complete phonological model should entertain the rich details encoded in the lexicon and an algebraic system extracted from the rich lexicon, and to this end a hybrid phonological framework is called for as concluded in Pierrehumbert (2006) and Guy (2007, 2014).

In fact, the phonological patterns that Halle (1985) refers to as evidence for lexical economy also justify the necessity of a rich lexicon other than an abstract grammar. First, even if we assume that English speakers can memorize the stress pattern of each word (e.g. to distinguish noun-verb (near) minimal pairs like pérfume vs. perfúme[9]), it does not imply that English speakers are unaware of the majority of antepenultimate stress. English speakers do not faithfully reproduce the stress in the Russian loanwords since the production of novel words should still conform to their native phonological generalization with a majority of antepenultimate stress patterns, which is common in loanword phonology.

Loanwords do not always obey native phonotactics, however. New phonemic contrasts can gradual emerge from nativizing foreign words that form (near-)minimal pairs with native words. This process is initiated by storing non-native words in a rich lexicon first, which eventually change the phonological generalization after gradually gaining a stronger lexical strength. For instance,

[9] cf. French speakers; see Pepperkemp & Dupoux (2002) and Pepperkemp et al. (2010).

previously the consonant /t/ in Japanese became [ts] before [u] and [tʃ] before [i] with no exception, and the process was applied to early loanwords as well (e.g. 'ticket' [tʃiketto]). However, more recently new loanwords such as 'party' have been adapted as [paati] rather than *[paatʃi] (Ito & Mester 1995). Bybee (2001:54) argues that each token of loanwords with non-native sounds is stored in the rich lexicon but at first the tokens have a very weak lexical strength and are produced without violating Japanese phonotactics. After perceiving and storing more foreign tokens with [ti], the lexical strength of these tokens is gradually reinforced as a pressure for these tokens to be produced faithfully. The contrast between /t/ and /tʃ/ is eventually established in Japanese. An abstract grammar and a rich lexicon are both required to explain the transition.

6. Probabilistic Selection of Input: Introduction

The above discussion in sum includes patterns that are compatible with a rich lexicon that stores surface paradigms faithfully with other lexical information such as token frequency to determine basic allomorphs at any given point of the acquisition course. Then, the crux of the matter is how to determine basic allomorphs. I propose that at any given point, each allomorph is assigned a probability between 0 and 1 for the allomorph to be selected as the input of the corresponding morpheme, which is thus referred to as Probabilistic Selection of Input (PSI). More specifically, each surface allomorph of a morpheme is possibly selected as the phonological input of the morpheme, and therefore the outcome of morphophonological acquisition is a set of input probabilities of surface allomorphs. By gradually gaining a dominant probability, a stored paradigm is conceived as a more basic allomorph compared to its competitors.

This approach has been recently proposed in Guy (2007, 2014) to explain the -t/d deletion patterns of the conjunction word 'and'. The coronal deletion process is well-known to be applied at a higher rate to words with a higher token frequency (referred to as 'exceptional words'), like 'and', as shown by the significantly different deletion rates in Table 1.1 (see also Neu 1980, Bybee 2002). Contexts are the other factor that affects the application of coronal deletion: An obstruent triggers the deletion of a preceding -t/d more easily, whereas in a prevocalic position, the deletion process is blocked more often.

	Non-exceptional	Exceptional ('and')
Pre-obstruent	58.3%	87.9%
Pre-vocalic	10.4%	75.3%
Difference	47.9%	12.6%

Table 1.1. Proportions of -t/d deletion of non-exceptional words and the exceptional word 'and' (18 speakers from the ONZE corpus, University of Canterbury; Guy 2014)

In an exceptional rule approach, we can simply assume an exceptional -t/d deletion rule that is applied only to 'and' more frequently. However, this exceptional rule approach cannot explain the intriguing interaction between context and exceptionality in Table 1.1: If the phonological definition of the regular and exceptional coronal deletion rules is identical (i.e. delete -t/d more before an obstruent), why is the contextual difference larger for non-exceptional words (47.9% vs. 12.6%)? Guy (2007, 2014) alternatively proposes a rich-lexicon approach (see also van Oostendorp 2014 and van de Weijer 2012, 2014) in which exceptional words have multiple URs, including the reduced form [æn] or [n], and the non-reduced form [ænd]. When the non-reduced variant /ænd/ is selected as the input, the exceptional word, like other non-exceptional words, undergoes the same deletion rule, which applies depending on the context. The reduced variant /æn/ or /n/, however, may also be selected across the contexts as the input and thus surface faithfully without a word-final [-d]. Consequently, the deletion rates of 'and' are less affected by different contexts since the variable deletion process has no effect when the reduced variants are selected as the input and always surface without [-d].[10]

The goal of this thesis is thus to build a morphophonological learning model based on this rich-lexicon approach, test its precision in modeling morphophonological acquisition and diachronic morphophonemic changes, and further contribute to the body of evidence for a rich lexicon.

[10] For example, assume that the probability for the deletion rule to apply is 50% before an obstruent and 10% before a vowel. A non-exceptional word with only one single UR with [-d] will always undergo the same deletion process and thus surface without [-d] 50% of the time before an obstruent but only 10% of the time before a vowel (i.e. a contextual difference of 40%). An exceptional word 'and', on the other hand, has two inputs /æn/ and /ænd/ which are selected presumably with an equal chance. Its output will surface without [-d] 75% of the time before an obstruent (i.e. 50% from /æn+C/→[æn+C] plus 50% × 50% = 25% from /ænd+C/→[æn+C]) and 55% of the time before a vowel (i.e. 50% from /æn+V/→[æn+V] plus 50% × 10% = 5% from /ænd+V/→[æn+V]) – a smaller contextual difference of 20%. See also Guy (2007) for a formal mathematical proof.

7. Probabilistic Selection of Input and constraint grammar

With the emphasis on the necessity of abstract phonological grammar in §5, its form and organization must be fully defined to complete a hybrid model. This dissertation adopts a constraint-based grammar, namely Optimality Theory (OT; Prince & Smolensky 1993/2004), to capture grammar learning in the current study of morphophonological acquisition. In Standard OT, outputs are not derived through a set of rule applications from input; rather, a set of violable innate constraints are ranked to evaluate the best output candidate of an input, which violates fewest and lower-ranked constraints. There are two conflicting types of innate constraints in Standard OT: Markedness constraints are violated by output candidates with phonologically marked segments or structures, and faithfulness constraints are violated by an output that deviates from its input (e.g. via deletion, epenthesis, feature changes, etc.). When the former are ranked higher, speakers choose outputs in which marked structures are avoided by sacrificing input-output consistency. On the other hand, dominant faithfulness constraints preserve input structures in the output to allow the emergence of marked elements. Language-specific phonological patterns are considered the product of language-specific constraint rankings, and phonological acquisition is a process of acquiring language-specific constraint rankings from a presumably universal initial state. Of our primary interest is this gradual learning process in the current study of morphophonological acquisition, which will be modeled by the Gradual Learning Algorithm developed in Boersma (1998) and Boersma & Hayes (2001). More details are spelled out in §2 of Ch. 2.

8. Design guidelines of a learning algorithm

The PSI model will be constructed by following the guidelines of designing a formal acquisition model as provided in Yang (2003). The first guideline is <u>formal sufficiency</u> – "the acquisition model must be causal and concrete" (2003:5); that is, the model should be capable of explaining how a learning outcome is achieved via transparent processes confirmed by mathematical proof or computer simulation. The input selection process satisfies this criterion as the calculation of the selection probability of each stored allomorph and its gradual changes will be defined and processed transparently. In terms of grammar learning, the adoption of the Gradual Learning Algorithm in §7 satisfies the requirement with a transparent re-ranking process represented by gradual numeric adjustments. Joining together, the two learning mechanisms will determine the phonological output of a morphologically complex form, whose learning progress can be monitored in computer simulation.

Second, the acquisition model must demonstrate explanatory continuity. Assuming that children and adults have the same linguistic competence, the structure of the grammar does not undergo any transformation along the course of language acquisition. Instead, learners only change grammar settings to approximate those of a target grammar. Standard OT states that constraints are innate and universal, which form a closed constraint set and therefore follow a narrower definition of continuity: Children and adults have the same set of innate constraints. In the proposed model, however, learnable constraints can be induced and added to an open constraint set, so continuity here requires a broader definition: The phonological grammar should be represented only by either innate or learned constraints, and for simplicity reasons in the model proposed in this thesis, I assume that only unnatural markedness can be invented by learners. Since no other forms of phonological grammar (e.g. rules) can be introduced in constraint-based theories by definition, the structure of the grammar remains unchanged through the learning process. Readers are referred to Appendix A for all constraints introduced in the rest of this thesis. In terms of lexicon, we shall assume that after morphological awareness, the same lexical factors are used by learners of different generations in the same input selection process to determine how basic allomorphs are selected in various stage of morphophonological acquisition.

The two criteria above focus on the internal structure of an acquisition model, and the third criterion – developmental compatibility – requires a correct learning outcome generated by the model at successive learning stages. It is not surprising that the learners switch back and forth between grammar assumptions and have different lexical constructions (e.g. lexicon size, word category, etc.) before the learning process reaches the end state. These changes are then reflected in the learners' performance as different production patterns, which should be captured by the model. Moreover, as required by the first criterion, the sequential variations must be explained by transparent processes. In Ch. 3, simulated results will be compared with the real output data collected at different stages of Dutch morphophonological acquisition, and the proposed learning model can be verified if no major mismatches can be found in different phases and an adult-like lexical generalization and constraint grammar are acquired.

Finally, I would like to add another criterion which I believe to be highly important as well. If the claim that most diachronic phonological changes are simply the results of language acquisition is accepted (e.g. Andersen 1973, 1978; Weerman 1993), we expect a learning model to demonstrate divergent acceptability. That is, since the learning data may not be evenly distributed (e.g. Zipfian distribution; Zipf 1949), it is not always the case that learners can receive learning inputs required for the acquisition of a target grammar from speakers of previous generations. Instead of rejecting any 'non-target' grammar, however, the learners should accept any generalization even if it may lead to some performance divergence from speakers of previous generations. This is the onset of a diachronic change, and the diverged patterns could stabilize over time and become regular patterns. We have seen the examples of rule inversion triggered by relexicalization above (i.e. accept a different allomorph as a more basic allomorph), and the proposed computational model should be able to simulate these gradual historical changes.

Only when all of the above four criterion are followed, an acquisition model will have enough explanatory power to help us gain insight into child language development. The proposed model will be designed in accord with the blueprint above, and aims at solving some perplexing morphophonological learning puzzles.

9. Summary

Following the foregoing discussion, it is important to further note that in this thesis, PSI is only among one of many ways to formalize how learners freely access a rich lexicon and how information retrieved from a rich lexicon interacts with grammar. The implementation of such an algorithm does not suggest that PSI is the single universal lexical process since it is easily undermined in many other circumstances as we will see toward the end of the thesis. Rather, the spirit inside the proposed model is that a rich lexicon must be a component that is universally available to all language learners, as suggested by the cross-linguistic evidence. Universal Grammar provides a mechanism that allows learners to determine how to efficiently use the abundant information stored in their rich lexicon and formulate or adjust their grammar to acquire a target language successfully. PSI is designed as a simple and computationally feasible algorithm, and is thus assumed as one of the strategies that learners of certain languages discussed in this thesis (see below) would invent and apply under the permission of Universal Grammar (cf. SURA). With PSI, it is then possible to at least partially explain how learners behave in acquiring adult-like morphophonological patterns, either successfully or unsuccessfully, in these particular languages.

The thesis is organized as follows. The proposed PSI learning algorithm will be developed in Ch. 2, including the details of lexical factors that can jointly determine the selection probability of stored allomorphs and the settings of a constraint that can be gradually adjusted. The formal comparison between a simulated output and real performance data in Dutch morphophonological acquisition will be arranged in Ch. 3. The same procedure will also be used in Ch. 4 and Ch. 5 to compare simulation results with experimental data collected to model surface variations after relexicalization in Mandarin and Korean. Finally, residual issues of the proposed model and a possible extension will be discussed in Ch. 6.

Chapter 2

Probabilistic Selection of Input with constraint grammar

This chapter seeks to build a complete structure of Probabilistic Selection of Input (henceforth PSI) and illustrates how it works. In §1 and §2, the two core elements of the proposed PSI model – a rich lexicon and a constraint grammar – will be spelled out: The former stores rich lexical information, from which the selection probability of stored surface allomorphs is computed. The latter captures grammatical generalization. §3 is a review of a widely accepted two-stage morphophonological learning course, which is introduced into the PSI model. In §4, an example is used to demonstrate how PSI works to update lexical information and adjust constraint ranking in the two independent stages. A simulation with toy data in §5 demonstrates different outcomes of grammar learning with various input distributions in PSI. In particular, it will be shown that PSI may not converge on a target grammar if an input distribution is highly skewed. Finally in §6, a comparison between PSI and other existing morphophonological acquisition models explains differences and advantages in the current proposal.

1. Selection probability of lexically stored allomorphs

The definition of a rich storage process in morphophonological acquisition in §1 of Ch. 1 is repeated here: Whenever learners perceive different allomorphs of a morpheme, the allomorphs are stored in the lexicon and become potential input candidates of the morpheme. Each stored allomorph of the morpheme is assigned a selection probability (SP) between 0 and 1 to be the input of the morpheme. In the hypothetical language constructed in Ch. 1 (repeated as (1) below), the input of 'Stem A (sg.)' can be either /bap/ or /bab/, and the input of 'Stem A (pl.)' can be either /bap+a/ or /bab+a/; whenever Stem-A is produced, the two allomorphs always compete against each other to be its input regardless of the context. The issue at stake is what lexical factors are involved in determining the SP of each stored allomorph, which will be addressed from §1.1 to §1.4.

(1) Morphophonemic alternations in hypothetical language

[bap]	'Stem-A (sg.)'	[ba.b-a]	'Stem-A (pl.)'
[bat]	'Stem-B (sg.)'	[ba.t-a]	'Stem-B (pl.)'

1.1 Lexical factor I: Token frequency

The important role of token frequency in (morpho-)phonology has been briefly summarized in §4 of Ch. 1, and this subsection aims to formalize token frequency as one of the lexical variables in the calculation of SPs in PSI.

In each learning cycle PSI, the algorithm feeds the 'machine learner' with one learning input. The token frequency of the learning input and the surface allomorph included in the input increases by one. After the frequency update process, the SP of an allomorph can be calculated with the latest token frequency information in this given learning cycle. The relation between token frequencies and SPs is straightforward: All else being equal, if an allomorph has a higher token frequency than its competitors, it also has a higher SP.[11] The frequency-based SP of an allomorph can be formalized as formula (2).

(2) By the definition of selection probability (SP), the SP of some allomorph *A* of morpheme *M*, relative to other allomorphs of *M*, is calculated as follows, where freq(A) is the token frequency of *A*, and freq(M) is the token frequency of *M.*

$$SP(A) = \frac{freq(A)}{freq(M)}$$

$$freq \geq 0, \qquad 0 \leq SP \leq 1$$

To an extreme, if the token frequency of a surface allomorph is much higher than that of its competitors, it is likely to be assigned an SP close to 1, and the allomorph is practically equal to the single UR of a morpheme; the allomorph would be almost always selected as the input of the morpheme, as if there is only one single input for the morpheme. However, since other low-frequency input options may still be selected as the input despite an extremely low chance, PSI is still fundamentally different from SURA models that store only one single UR for each

[11] See Ch. 3 for the similarity in the frequency distribution of morphological forms between adult and child corpora. Beyond the scope of morphophonological acquisition, Zamuner et al. (2005) also found that English-learning children tend to produce coda consonants which occur in the ambient input (e.g. adult speech) more frequently, rather than "conform to what is unmarked across languages" (e.g. coronals).

morpheme. The only possibility for PSI to assign a stored allomorph an SP of 1 and thus literally exclude other low-frequency allomorphs from the input selection process is that the low-frequency allomorphs completely fade away as memory decays (see §1.4).

The correlation between token frequencies and SPs in PSI parallels Albright's (2002, 2008, 2010) Single Surface Base Restriction which assumes that only surface-true allomorphs can be the base (or basic form) of a morpheme. In PSI, it is not necessary to stipulate such a restriction since non-surface-true paradigms must have a zero frequency and cannot be stored in the lexicon as a possible input of a morpheme.[12] Another advantage of storing only surface-true allomorphs is the ease of computation as discussed in §1 of Ch.1: Limited input choices allows a considerably reduced computation space. In this light, I assume that learners obey such a 'restriction' to store and access the observable surface paradigms to optimize lexical processing and facilitate morphophonological acquisition.

Some token frequency effects have been addressed in some learning models, but mostly they emerge as a by-product of the learning process in these models, rather than serve as a lexical variable that guides learners toward any decision on selecting the input of a morpheme. See §6 for a more detailed discussion of various morphophonological learning models.

1.2 Lexical factor II: Individual Error Proportion

We have seen in the previous sections that plural allomorphs in the hypothetical language should be assigned a dominant SP to converge on the correct grammar, and to this end, plural allomorphs should have a dominant token frequency if PSI only contains the formula (2). However, high-frequency allomorphs may not always be the allomorphs that can lead to a correct grammar generalization. For example, since singular forms tend to occur more frequently than plural forms (70~85% vs. 15~25%, Greenberg 1966:32; see Ch. 3 for a similar distribution in Dutch),[13] the allomorph /bap/ of Stem-A in the hypothetical language is expected to have a higher token frequency than /bab/. In this case, the input of Stem-A (pl.) has a higher

[12] I will not discuss lexical representations built from misperception (e.g. Macken 1980) or templatic lexical representations (e.g. Ferguson & Farwell 1975) as a consequence of insufficient exposure to learning targets in this dissertation, but I agree with the possibility of gaining input candidates from multiple sources other than the auditory channel, such as orthography (see Ch. 6).

[13] This is possibly because "humans tend to focus on individuals (and to treat groups as individuals, e.g. *herd*, *battalion*, *cloud*" (Haspelmath 2006:45).

chance to be /bap+a/. With a grammar that requires a faithful input-output mapping, an output error *[ba.p-a] may surface from /bap+a/. In sum, the privilege of more frequent forms should be acknowledged, but since the universal goal of language acquisition is to produce more correct outputs than output errors to facilitate communication processes, allomorphs which have a higher chance to generate more target forms should also be recognized as better input options in PSI. I thus assume that each allomorph is also evaluated with its own chance of generating output errors (cf. §1.3).

The second lexical factor, **Individual Error Proportion** (IEP) is thus proposed for each stored allomorph, which is directly related to the number of output errors generated by selecting an allomorph as the input. The IEP of an allomorph *A* can be calculated using formula (3) by tracking the number of output errors and correct outputs produced with any input that contains *A* regardless of morphological environments. In every learning cycle, IEP(A) is updated immediately after an output error or a correct output surfaces by selecting *A* as the input. If output errors are far fewer than correct outputs, a lower IEP(A) is generated to reward *A*'s higher efficiency of producing correct outputs.

(3) The Individual Error Proportion (IEP) for some allomorph *A* of morpheme *M* is defined as follows, where Error(A) is the number of output errors generated with *A*, and Correct(A) is the number of correct outputs generated with *A*.

$$IEP(A) \equiv \frac{Error(A)}{Error(A) + Correct(A)}$$

$$Error \geq 0, \qquad Correct \geq 0, \qquad 0 \leq IEP \leq 1$$

The IEP of the two allomorphs of Stem-A in the toy data can be calculated in (4): Assume that the two allomorphs are both previously selected 60 times in the singular context and 60 times in the plural context. With a grammar that strictly forbids intervocalic voicing, the 60 times of selecting /bap+a/ as the input of 'Stem-A (pl.)' lead to 60 output errors *[ba.p-a], whereas selecting /bap/ as the input of 'Stem-A (sg.)' can always surface faithfully as the correct output [bap] (i.e. $\text{Correct}_{/\text{bap}/}$ = 60). When /bab+a/ is selected as the input of 'Stem-A (pl.)', it can always surface faithfully as the correct output [ba.b-a] without undergoing intervocalic voicing (i.e. $\text{Correct}_{/\text{bab}/}$ = 60). In the singular context, it is assumed that the grammar may sometimes incorrectly allow vowel epenthesis or deletion to repair a voiced obstruent coda (i.e. /bab/→*[ba.bi] or /bab/→*[ba]; see §5) and that only 40 out of 60 selections of /bab/ as the input lead to the correct output [bap] via final devoicing. Nevertheless, the plural allomorph still surfaces as more correct outputs across contexts than the singular allomorph (i.e. $Correct_{/bab/}$ = 60 + 40 = 100 > $Correct_{/bap/}$ = 60), which gives rise to a much lower $IEP_{/bab/}$.

(4) Calculating the $IEP_{/bab/}$ and $IEP_{/bap/}$

$Error_{/bab/}$ = 20, $Correct_{/bab/}$ = 100, $IEP_{/bab/}$ = 20 / (20 + 100) = 0.17

$Error_{/bap/}$ = 60, $Correct_{/bap/}$ = 60, $IEP_{/bap/}$ = 60 / (60 + 60) = 0.5

Token frequency is the base variable for every surface allomorph, which is adjusted with other lexical variables like IEP, and the product of the adjustment is called Adjusted Token Frequency (ATF). The ATF of each allomorph can be computed by multiplying the token frequency with the corresponding IEP to the power of -1 as shown in the formula (5), and IEP is thus inversely related to ATF; a raw token frequency thus grows significantly with an extremely low IEP, whereas an ATF and a raw token frequency are identical with an IEP is 1 (i.e. 100% error rate). That is, allomorphs which generate fewer errors are promoted with a higher ATF as noted earlier.

(5) The Adjusted Token Frequency (ATF) of some allomorph A is defined as follows, where freq(A) is the raw token frequency of some allomorph A and IEP(A) is the Individual Error Proportion of *A.*

$$ATF(A) \equiv freq(A) \times IEP(A)^{-1}$$

$$freq \geq 0, \qquad ATF \ \geq 0, \qquad 0 \leq IEP \leq 1$$

Assuming that the singular allomorph /bap/ has a higher token frequency than the plural allomorph /bab/, the ATF of the two allomorphs can be calculated in (6). Although $freq_{/bap/}$ is twice higher, $ATF_{/bap/}$ is slightly lower than $ATF_{/bab/}$ for a higher error proportion.

(6) Calculating the $ATF_{/bab/}$ and $ATF_{/bap/}$

$freq_{/bab/}$ = 200, $ATF_{/bab/} = 200 \times (0.17)^{-1}$ = 1176.5

$freq_{/bap/}$ = 400, $ATF_{/bap/} = 400 \times (0.5)^{-1}$ = 800

The SP calculation can be revised as the formula in (7) below.

(7) By the definition of selection probability (SP), the SP of some allomorph *A* of morpheme *M*, relative to other allomorphs of *M*, is calculated as follows, where ATF(A) is the Adjusted Token Frequency of *A*, and ATF(x) is the Adjusted Token Frequency of the x^{th} allomorph of *n* allomorphs of *M.*

$$SP(A) = \frac{ATF(A)}{\sum_{x \in M}^{n} ATF(x)}, 0 \leq SP \leq 1$$

With ATF, the SP of an allomorph *A* is equal to the proportion of *A*'s ATF to the sum of the ATF of *A* and all of its competitors. The above calculation of IEP and ATF then allows us to derive the SP of each allomorph as in (8).

(8) Calculating the $SP_{/bab/}$ and $SP_{/bap/}$

$SP_{/bab/}$ = 1176.5 / (1176.5 + 800) = 0.595

$SP_{/bap/}$ = 800 / (1176.5 + 800) = 0.405

Now consider another scenario: Coincidentally, the number of output errors is zero for the two allomorphs of Stem-A, which might occur at the very beginning of morphophonological acquisition. With a zero numerator, the IEP of the both allomorphs is naturally zero, and consequently the corresponding ATFs will be zero as well by dividing a raw token frequency with a zero IEP. Eventually, the SP of either allomorph is infinity with a fraction of 0 / 0 and results in a computational problem. The solution is to apply smoothing by adding 0.5 to the numerator and 1 to the denominator of the IEP calculation, which changes the IEP formula from (3) to (9). The influence from these small numbers on IEP is expected to decline as the number of errors and correct outputs grows.

(9) The Individual Error Proportion for some allomorph *A* of morpheme *M* is defined as follows, where Error(A) is the number of output errors generated with *A*, and Correct(A) is the number of correct outputs generated with *A*.

$$IEP(A) \equiv \frac{Error(A) + 0.5}{Error(A) + Correct(A) + 1}$$

$$Error \geq 0, \quad Correct \geq 0, \quad 0 < IEP \leq 1$$

One important property of the above IEP calculation is the intrinsically underestimated contribution from targets to IEP; when correct outputs outnumber errors, the amount changed in IEP by each output error is higher than by each correct output. For example, assuming that IEP(A) is (1 + 0.5) / (1 + 2 + 1) = 0.375, adding an output error raises IEP(A) to (2 + 0.5) / (2 + 2 +1) = 0.5 (i.e. difference = 0.5 - 0.375 = 0.125). Adding another correct output, however, only lowers IEP(A) to (1 + 0.5) / (1 + 3 + 1) = 0.3 (i.e. difference = 0.375 - 0.3 = 0.075). In other words, adding an error greatly raises IEP, but adding a correct output only slightly decreases IEP.[14] Such an asymmetry is necessary to avoid overestimating the SP of 'worse' allomorphs which at times produce correct outputs with different grammar assumptions (e.g. /bap+a/→[ba.b-a] with an intervocalic voicing grammar). I will return to this case below.

It is also possible to convert the formula (7) to (10) to predict the required IEP(A) for a specific SP(A).

(10) The IEP(A) of a specific SP of some allomorph *A* of morpheme *M* is calculated as follows, where *M'* is a set of allomorphs of *M* excluding *A*, and ATF(x) is the Adjusted Token Frequency of the x^{th} allomorph of *n* allomorphs of *M'*.

$$IEP(A) = \frac{(1 - SP(A)) \times freq(A)}{SP(A) \times \sum_{x \in M'}^{n} ATF(x)}$$

When the allomorphs /bab/ and /bap/ of Stem-A has a frequency of 200 and 400 respectively as assumed above and an $IEP_{/bap/}$ of 0.4, the $IEP_{/bab/}$ for an $SP_{/bab/}$ of 0.9 must be as low as 0.022 to compensate for the lower $freq_{/bab/}$ in (11). That is, for each output error generated with the stem input /bab/, nearly 50 target outputs must be generated as well to compensate for the output error to achieve the low $IEP_{/bab/}$. This transparent relation between lexical variables later allows us to explore why allomorphs with different frequency distributions can or cannot be assigned a dominant SP.

[14] This asymmetry in adjusting IEPs is reversed when output errors outnumber correct outputs; IEP increases slightly by adding an output error but decreases more significantly when adding one correct output. However, with more output errors, an allomorph might be reliably identified as a worse input option, and it is no longer necessary to overestimate the effect of an output error in this case.

(11) Calculating the $IEP_{/bab/}$ for an $SP_{/bab/}$ of 0.9

$$IEP_{/bab/} = \frac{(1-0.9)\times 200}{0.9\times\frac{300}{0.4}} = 0.022$$

Finally, such individual output error proportions account for another type of token frequency effects: Morphologically complex forms with a high token frequency are acquired at an earlier stage since allomorphs of a morpheme contained in these forms can be 'tested' more often; the number of output errors generated with allomorphs in these forms accumulates faster for learners to identify 'worse' input options more rapidly. For example, if Stem-A (pl.) is produced at a higher rate than other plural forms, the input with the stem's singular allomorph (i.e. /bap+a/) will be selected as the plural form's input more often than the same input option of other plural forms. Consequently, the input /bap+a/ may surface as more output errors (i.e. *[bap-a]) at any given point, and the singular allomorph of Stem-A is recognized as a worse input choice prior to the singular allomorph of other stems. This prediction is also borne out in Jarosz's (2011) modeling on the acquisition of Dutch voicing alternation with a rich lexicon, in which a stem-final /d/ in plural forms is produced less accurately due to the low token frequency of such patterns for learners to take longer to converge on the correct UR of these forms (see also §4.8 for a similar prediction in Albright (2002, 2008, in press)). This developmental difference is also demonstrated in Kerkhoff's (2007) experimental results and will be verified with computer simulation in Ch. 3.

1.3 Lexical factor III: Morphological Error Proportion

The above SP calculations suggest that allomorphs from the same morphosyntactic context are fully independent from each other since their raw token frequency and the number of output errors and correct outputs only contribute to their own SP. Thus, although in theory all plural allomorphs in the hypothetical language should be recognized as a 'better' input since they can generate fewer output errors across morphosyntactic contexts, independency predicts that some plural allomorphs are much 'better' than others at a certain point, given that they are 'tested' by the learners with a different frequency (see §1.2). While the SP variation across allomorphs from the same context should be empirically true and are intended to be modeled in PSI, we might expect learners to extend the success of one allomorph to promote another if they are extracted from the same morphosyntactic context. For instance, if plural allomorphs can overall generate more correct outputs than errors, the entire set of plural allomorphs should have a privilege in the input selection process (cf. Albright 2002, et seq.).

Morphological privilege demonstrated by selecting a 'basic' morphological class for paradigm uniformity is attested in both early language acquisition and diachronic changes. In their longitudinal study of four Portuguese-learning children's production, Simões & Stoel-Gammon (1979) reported that verb stems were leveled to their third personal singular allomorph in the first person singular context. Stem vowels in Portuguese change along with different person marker as shown in (12). With the first person singular marker [o], the stem vowels are raised as [o], [e] or [i], [u] in the two different conjugations but lowered unanimously as [ɔ] and [ɜ] in the third person singular context marked with the suffix [e]. In child speech, however, the vowel alternation does not apply in the first person singular context even if the inflection suffix [o] is correctly attached, and the stem vowels are leveled to those in the third person singular paradigms. Hale (1973) also found similar evidence that Maori passive forms are re-analyzed by learners so roots are leveled to underived forms.

(12) Third singular dominance in the Portuguese child language (cited in Hooper 1980:167)

	2nd conjugation		3rd conjugation	
infinitive	'eat'	'drink'	'sleep'	'get'
first sg.	c[o]mo	b[e]bo	d[u]rmo	cons[i]go
third sg.	c[ɔ]me	b[ɜ]be	d[ɔ]rme	cons[ɜ]ge
first sg. (child)	c[ɔ]mo	b[ɜ]bo	d[ɔ]rmo	cons[ɜ]go

As reviewed in §4.3 of Ch. 1, child-speech patterns has been frequently observed in diachronic changes, and paradigmatic leveling demonstrated in child morphophonology is also attested in diachronic sound changes. In Middle High German, stem vowels alternated with different person markers as in (13a), but when Yiddish emerged from Middle High German, stem vowels were leveled to the first singular allomorph in (13b). Paradigmatic leveling in Yiddish is atypical not only because it is unattested in other related dialects, but it also contradicts the observation that third person singular allomorph is usually the most basic paradigm (e.g. Bybee 1985, and see the above Portuguese example) and is acquired earlier than any other persons (e.g. Bates 1976, Hooper 1980).

Albright (2002:7) claims that a basic paradigm must be 'maximally informative' to avoid (i) phonological and morphological neutralization and (ii) exceptions to the morphophonological rules that derive the surface forms from the basic paradigm. The first person paradigms in Yiddish, but not in Middle High German, satisfied the above criteria, which were thus considered the basic paradigms in Yiddish and led to paradigmatic leveling.[15]

(13) Stem vowel alternation/leveling in Middle High German and Yiddish (Albright 2002:17)[16]

a. Middle High German

'dig'	sg.	pl.	'know'	sg.	pl.
1st	grɑbe	grɑben	1st	w**ei**ʒ	weiʒʒen
2nd	gr**e**best	grabet	2nd	w**ei**st	weiʒʒet
3rd	gr**e**bet	graben	3rd	w**ei**ʒ	weiʒʒen

b. Yiddish

'dig'	sg.	pl.	'know'	sg.	pl.
1st	grɔb	grɔbən	1st	v**ey**s	v**ey**sən
2nd	grɔbst	grɔbt	2nd	v**ey**st	v**ey**st
3rd	grɔbt	grɔbən	3rd	v**ey**st	v**ey**sən

The current PSI model is incapable of capturing any morphological privilege since changes in the lexical variables of their corresponding allomorph do not affect the state of other allomorphs from the same morphosyntactic context. Consider the

[15] Albright (2002:22) only found three exceptions to the 1sg paradigm leveling pattern: zany(ən) 'to be' has not undergone leveling, an auxiliary verb (i.e. velən) is not derived from 1sg present indicative, and gefelən 'be pleasing' is used predominantly in the 3rd person (thus is derived from a 3sg form). In particular, the first two, according to Albright, are verb paradigms with an extremely high frequency, which is the force against a general paradigm leveling tendency.

[16] As noted in fn. 1 in Albright (2002:17), the leveling patterns summarized here only hold true for the eastern Yiddish dialect spoken in Central and Eastern Europe; whether other dialects demonstrate the same patterns remains an open question.

examples in (14), where two pairs of stem allomorphs have the same singular-plural token frequency ratio (2:1). If, at the very beginning, $IEP_{SG\text{-}A}$ (i.e. Individual Error Proportion of the singular form of Stem-A) is equal to $IEP_{SG\text{-}B}$ and $IEP_{PL\text{-}A}$ is also identical to $IEP_{PL\text{-}B}$, $SP_{PL\text{-}A}$ and $SP_{PL\text{-}B}$ are indistinguishable.

(14) Same initial $SP_{PL\text{-}A}$ and $SP_{PL\text{-}B}$

$$SP(A) = (freq(A) \times IEP(A)^{-1})/(freq(A) \times IEP(A)^{-1} + freq(B) \times IEP(B)^{-1})$$

Pair A: $SG_A \sim PL_A$	$freq_{SG\text{-}A} = 2$, $freq_{PL\text{-}A} = 1$	$IEP_{SG\text{-}A} = IEP_{SG\text{-}B} = 0.5$
Pair B: $SG_B \sim PL_B$	$freq_{SG\text{-}B} = 2$, $freq_{PL\text{-}B} = 1$	$IEP_{PL\text{-}A} = IEP_{PL\text{-}B} = 0.2$

$SP_{PL\text{-}A} = (1 \times 0.2^{-1}) / (2 \times 0.5^{-1} + 1 \times 0.2^{-1}) = 0.556$

$SP_{PL\text{-}B} = (1 \times 0.2^{-1}) / (2 \times 0.5^{-1} + 1 \times 0.2^{-1}) = 0.556$

Now assume that the stem *A* is much more frequent than *B* and is produced in the next few learning cycles, in which the plural allomorph of *A* surfaces as correct outputs. $IEP_{PL\text{-}A}$ is thus lower to, say, 0.05 as in (15). The immediate consequence is that PL_A is recognized as a much better input option of *A* with a substantially raised $SP_{PL\text{-}A}$, whereas the status of PL_B remains unchanged.

(15) Lowered $IEP_{PL\text{-}A}$ does not affect $SP_{PL\text{-}B}$

$$SP(A) = (freq(A) \times IEP(A)^{-1})/(freq(A) \times IEP(A)^{-1} + freq(B) \times IEP(B)^{-1})$$

$SP_{PL\text{-}A} = (1 \times 0.05^{-1}) / (2 \times 0.5^{-1} + 1 \times 0.05^{-1}) = 0.833$

$SP_{PL\text{-}B} = (1 \times 0.2^{-1}) / (2 \times 0.5^{-1} + 1 \times 0.2^{-1}) = 0.556$

For the emergence of a plural morphological privilege as shown in Albright's base identification approach (see also Albright & Hayes 2003 and §4.8 of Ch. 5), an additional lexical parameter referring to the entire morphosyntactic context must be included in PSI, which promotes PL_B and assign it a higher SP as more correct outputs are generated from PL_A, despite the fact that PL_B does not contribute to those correct outputs.

To this end, I propose **Morphological Error Proportion** (MEP) which tracks the number of output errors and correct outputs across all morphological contexts generated from all allomorphs extracted from the same morphosyntactic context. MEP is updated along with IEP right after an output error or correct output is generated in each learning cycle. In the hypothetical language example, the context $[X+\emptyset]_{SG}$ can be extracted from singular allomorphs (i.e. stem plus a vacuous affix as a singular form) and $[X+a]_{PL}$ from all plural allomorphs (i.e stem plus /a/ as a plural form).[17] $MEP[X+\emptyset]_{SG}$ and $MEP[X+a]_{PL}$ can thus be formally defined in (16): The sum of output errors generated from the allomorphs from the same context divided by the sum of these allomorphs' output errors and correct outputs.

(16) The Morphological Error Proportion for the allomorphs from the context $[X+\emptyset]_{SG}$ and $[X+a]_{PL}$ is defined as follows, where *x* is the x^{th} allomorph of *n* allomorphs of the same context (i.e. $[X+\emptyset]_{SG}$ or $[X+a]_{PL}$), Error(x) is the number of errors generated by x^{th} allomorphs of the *n* allomorphs from the same context, and Correct(x) is the number of correct outputs generated by the x^{th} allomorphs of the *n* allomorphs from the same context.

$$MEP[X+\emptyset]_{SG} \equiv \frac{\sum_{x\in[X+\emptyset]SG}^{n} Error(x) + 0.5}{\sum_{x\in[X+\emptyset]SG}^{n} Error(x) + \sum_{x\in[X+\emptyset]SG}^{n} Correct(x) + 1}$$

$$MEP[X+a]_{PL} \equiv \frac{\sum_{x\in[X+a]PL}^{n} Error(x) + 0.5}{\sum_{x\in[X+a]PL}^{n} Error(x) + \sum_{x\in[X+a]PL}^{n} Correct(x) + 1}$$

$$0 \leq MEP \leq 1$$

Adding MEP to the current model changes the calculation of Adjusted Token Frequency (ATF) to formula (17), in which a raw token frequency are multiplied with the product of an IEP and MEP to the power of -1. The SP formula (7) remains unchanged and is repeated as (18) below.

(17) The Adjusted Token Frequency (ATF) of some allomorph *A* from context *C* of morpheme *M* is defined as follows, where freq(A) is the raw token frequency of some allomorph *A* and IEP(A) is the Individual Error Proportion of *A*, and MEP(C) is the Morphological Error Proportion of *C*.

$$ATF(A) \equiv freq(A) \times (IEP(A) \times MEP(C))^{-1}$$

[17] This extraction process does not only apply to concatenative morphology in which an affix is attached to a stem. For example, it is possible to extract ablaut patterns such as $[X\Lambda N]_{PASTPar}$ from irregular English past participle verbs such as *swum*, *run*, *begun*, etc. (see Bybee & Moder 1983), along with the regular context $[X+ed]_{PASTPar}$. Each context can have a corresponding MEP in the current proposal.

(18) By the definition of selection probability (SP), the SP of some allomorph *A* from context *C* of morpheme *M*, relative to other allomorphs of *M*, is calculated as follows, where ATF(A) is the Adjusted Token Frequency of *A*, and ATF(x) is the Adjusted Token Frequency of the x^{th} allomorph of *n* allomorphs of *M.*

$$SP(A) = \frac{ATF(A)}{\sum_{x \in M}^{n} ATF(x)}, 0 \leq SP \leq 1$$

Figure 2.1 represents an example of the morphological privilege in PSI in the hypothetical data. When the target of Stem-A (pl.) [ba.b-a] is derived from the input /bab+a/, $IEP_{/bab/}$ is lowered, and so is $MEP[X+a]_{PL}$, which jointly give rise to a higher $SP_{/bab/}$. Assuming two additional plural forms with a stem-final voiced obstruent such as [ba.z-a] and [ba.g-a], the IEP of the plural allomorphs /baz/ and /bag/ remains unchanged but the SP still increases as a global influence from the lowered $MEP[X+a]_{PL}$.

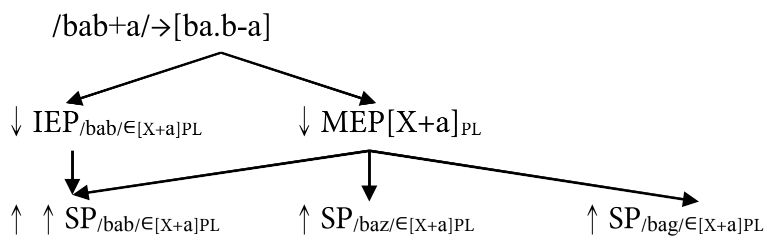


Figure 2.1. A global influence from MEP when producing a correct plural form with a plural allomorph

In (19), this example is illustrated with a presumed change of $IEP_{/bab/}$ and $MEP[X+a]_{PL}$ from 0.4 to 0.3 after a target [ba.b-a] is generated with the input /bab+a/ (bolded numbers). Before the changes, the three SPs are similar to each other. $SP_{/bab/}$ greatly increases after the changes since both $IEP_{/bab/}$ and $MEP[X+a]_{PL}$ drop to 0.3. The global influence from the change in $MEP[X+a]_{PL}$ can be observed in the increase $SP_{/baz/}$ and $SP_{/bag/}$, despite the same $IEP_{/baz/}$ and $IEP_{/bag/}$.

(19) A global influence on SP from MEP

$$SP(A) = (freq(A) \times (IEP(A) \times MEP(C))^{-1}) / (freq(A) \times (IEP(A) \times MEP(C))^{-1} + freq(B) \times (IEP(B) \times MEP(C'))^{-1})$$

Before /bab+a/→[ba.b-a]:

$SP_{/bab/} = (20*(0.4*0.4)^{-1})/(20*(0.4*0.4)^{-1}+50*(0.5*0.5)^{-1}) = 0.385$

$SP_{/baz/} = (10*(0.3*0.4)^{-1})/(10*(0.3*0.4)^{-1}+30*(0.4*0.5)^{-1}) = 0.357$

$SP_{/bag/} = (15*(0.5*0.4)^{-1})/(15*(0.5*0.4)^{-1}+40*(0.55*0.5)^{-1}) = 0.340$

After /bab+a/→[ba.b-a]:

$SP_{/bab/} = (20*(\mathbf{0.3}*\mathbf{0.3})^{-1})/(20*(\mathbf{0.3}*\mathbf{0.3})^{-1}+50*(0.5*0.5)^{-1}) = 0.526$ ↑ ↑

$SP_{/baz/} = (10*(0.3*\mathbf{0.3})^{-1})/(10*(0.3*\mathbf{0.3})^{-1}+30*(0.4*0.5)^{-1}) = 0.425$ ↑

$SP_{/bag/} = (15*(0.5*\mathbf{0.3})^{-1})/(15*(0.5*\mathbf{0.3})^{-1}+40*(0.55*0.5)^{-1}) = 0.407$ ↑

Likewise, in the Portuguese case, we can assume that $MEP[X+e]_{3rd.sg}$ is lower than $MEP[X+o]_{1st.sg}$ and thus increases the chance for third singular paradigms to be identified as better input options. In the Yiddish case, we can attribute the diachronic change to a gradually lowered $MEP[X+\emptyset]_{1st}$, which allows the 1st person paradigm to gain a dominant SP.

As in the previous section, we can predict the expected IEP or MEP required for a specific SP by alternating formula (10) as two formulas in (20) and (21). We will see how they help discover the reason why allomorphs with a lower frequency cannot be assigned a dominant SP in §2.5.

(20) An IEP for a specific SP of an allomorph *A* (i.e. *SP*(*A*)) from a morphological context *C* can be calculated as follows, where *M'* is a set of allomorphs of *M* excluding *A*, and *ATF(x)* is the Adjusted Token Frequency of the x^{th} allomorph of *n* allomorphs of *M'*.

$$IEP(A) = \frac{(1 - SP(A)) \times freq(A)}{SP(A) \times \sum_{x \in C \in M'}^{n} ATF(x)} \times \frac{1}{MEP(C)}$$

(21) An MEP for a specific SP of an allomorph *A* (i.e. *SP*(*A*)) from a morphological context *C* can be calculated as follows, where *M'* is a set of allomorphs of *M* excluding *A*, and *ATF(x)* is the Adjusted Token Frequency of the x^{th} allomorph of *n* allomorphs of *M'*.

$$MEP(C) = \frac{(1 - SP(A)) \times freq(A)}{SP(A) \times \sum_{x \in C \in M'}^{n} ATF(x)} \times \frac{1}{IEP(A)}$$

Like IEP, MEP is assumed to be unaffected by a stem morpheme without multiple allomorphs; e.g. producing either [bat] 'Stem-B (sg.)' or [ba.t-a] 'Stem-B (pl.)' as target outputs or errors does not shed light on whether the singular or plural class should be more dominant since the outputs are all derived from the same input of Stem-B.

Before moving on, possible concern over extending IEP as MEP might be their similar effect in simply different domains – IEP applies to individual allomorphs but MEP affects a set of allomorphs. As explained earlier in §1.2, IEP is necessary to capture individual SP differences caused by different numbers of output errors and speeds of accumulating errors varying across individual allomorphs. By contrast, MEP models the tendency shown by a set of allomorphs from the same morphosyntactic context. Both will be shown necessary to successfully model real learners' performance in morphophonological acquisition in Ch. 3.

Note that the proposal of this global effect does not imply that speakers will not be able to choose allomorphs from inconsistent morphological contexts as preferred inputs. In particular, token frequency might be higher for the allomorph *A* of a morpheme but for the allomorph *B* of another morpheme, where *A* and *B* are from different morphological contexts. In this case, a distributional difference might be stronger than the global effects as we will see in Ch. 5. In addition, the global effect might be absent when a consistent morphosyntactic context cannot be generalized by learners over a set of surface allomorphs. In such a case, MEP cannot contribute to the SP calculation of the allomorph set. For example, in Mandarin, the concave tone (i.e. 213, Tone 3) only surfaces faithfully in the phrase-final position, and contour simplification shortens the concave tone as a dipping tone (i.e. 21) in every non-phrase-final position, which is known as the Half-Sandhi rule (see Zhang & Lai 2010). A common morphosyntactic context of the dipping tone allomorphs cannot be specified as non-final positions may not represent any unified morphosyntactic context. The SP of each allotone of Tone 3 words is thus solely determined by token frequency and IEP. The modeling of the Mandarin diachronic tonal changes without MEP will be discussed in Ch. 4.

1.4 Lexical factor IV: Memory Decay

The experimental results in Goldinger (1996) summarized in §4.2 of Ch. 1 not only suggest that word tokens perceived previously are better recalled later if these tokens are produced by the same speakers, but also reveals the expected effect of memory decay: The recall rate drops significantly when the recall experiment is conducted a week later, if compared to the results in the same experiment conducted just five minutes later. Readers are referred to Kirchner (2011), Nosofsky (1988), Pierrehumbert (2001), Wedel (2006, 2007), and many others for the discussion about the decay effect in Exemplar-based models, and Becker & Tessier (2011) and Tessier (2007, 2009) for implementing a decay effect in Optimality Theory (OT).

Memory decay affects the selection process in PSI by eliminating input options; while /bab/ and /bap/ are both initially stored as possible input options of Stem-A, the former may have an extremely low token frequency and is thus more vulnerable to memory decay. In case that the allomorph /bab/ completely fades away from the lexicon, leaving the single input option /bap+a/ for Stem-A (pl.), the input surfaces as an output error *[ba.p-a] with a grammar that requires input-output faithfulness.

In the current PSI model, memory decay is introduced as a lexical factor which subtracts a fixed amount from the raw token frequency of each stored allomorph after each learning cycle. When the raw token frequency of an allomorph drops to zero due to memory decay, the allomorph completely fades away. If the allomorph is perceived once again later, its raw token frequency is reset to 1. The subtracted number is neither determined with any specific correlation between the number of learning cycles in PSI and real learning time spans nor is ideally calculated as a function of the corpus size, the mean token frequency and scattering of training inputs, etc. Instead, it is set rather arbitrarily from Ch. 3 to Ch. 5 following a prediction that allomorphs that can surface as correct outputs across morphological contexts but are perceived relatively infrequent are lexically unstable: When these allomorphs are absent from the lexicon, speakers are forced to choose other input options for corresponding morphemes and produce a great number of output errors in different contexts.

For example, in a mini-corpus of Dutch used as a learning input source in the simulation in Ch. 3, some ‘better’ allomorphs occur only two times or less per 1,000 tokens in the learning inputs, and the corresponding morphemes are produced with a great number of output errors. The error pattern is ascribed to the lower token frequency of the 'better' input option of these morphemes, which are more easily to decay from the lexicon and force speakers to choose a 'worse' input option. The decay rate is thus set to 0.002 in this Dutch case to approximate the error patterns. Similar predictions are referred to as an arbitrary criterion to set the decay rate in Ch. 4 and 5. The decay effect is only expected to significantly affect a small proportion of the learning targets, and the seemingly arbitrary decay rates should not undermine the goal and change the simulation outcome remarkably.

The calculation of Adjusted Token Frequency (ATF) in (17) can be modified to (22) to incorporate the memory decay effect. Note that since token frequency cannot be negative, the lower bound of token frequency must be zero.

(22) The ATF of some allomorph A from context C of morpheme M is defined as follows, where D is the decay rate and N is the number of learning cycles.

$$ATF(A) \equiv (freq(A) - D \times N) \times (IEP(A) + MEP(C))^{-1}$$
$$if\ (freq(A) - D \times N) < 0, freq(A) = 0$$

1.5 Summary of Probabilistic Selection of Input

The production and perception of a morpheme with multiple surface allomorphs in each learning cycle with the above lexical elements is summarized in (23) below.

(23) Producing and perceiving a morpheme with multiple surface allomorphs

a. When a surface allomorph of a morpheme is observed in a learning cycle, its raw token frequency increases by 1.
b. The SP of each stored allomorph of the morpheme is calculated with the allomorph's raw token frequency, IEP, and MEP shared with other allomorphs from the same morphosyntactic context in this given learning cycle.
c. One of the allomorphs is probabilistically selected as the input of the morpheme.
d. A correct output allows the algorithm to lower the IEP of the selected allomorph and the MEP of the allomorph's morpholosyntactic context. Otherwise, the algorithm raises the allomorph's IEP and MEP for an output error.
e. Each stored allomorph slightly decays from the memory by subtracting a fixed amount from the raw token frequency of the selected allomorph.

The crucial feature in this learning model is re-emphasized here: PSI does not guarantee assigning a dominant SP to a single allomorph that can always correctly surface as correct outputs since the algorithm does not seek a perfect solution to a 'UR problem' as in SURA models. Rather, the algorithm evaluates which allomorph has a greater chance to surface as adult-like outputs; if an allomorph can almost always surface as target outputs, it should be trusted and selected more often as input to mimic adults' performance. PSI may eventually assign a dominant SP to an allomorph that is also identified as the single UR in SURA models. This is because the allomorph that is expected to surface correctly across different contexts in SURA models (e.g. /bab/→[bap] and /bab+a/→[ba.b-a]) naturally has the greatest chance of being a target output in PSI, and is thus assigned a dominant SP. Nevertheless, PSI also evaluates the token frequency of an allomorph, which may strongly bias toward a 'less perfect' allomorph and eventually lead to a gradual shift of basic allomorphs.

2. Constraint grammar and Gradual Learning Algorithm

Previous sections have addressed how SPs can be computed with lexical variables by PSI learners, and changes in the selection of allomorphs must result in changes in the grammar and vice versa. As we have seen previously, the input of Stem-A (pl.) can be /bap+a/ if the allomorph /bap/ is accidentally recognized as a more basic allomorph of Stem-A at a given point. To derive the target output [ba.b-a] from /bap+a/, a short-term or permanent grammar shift to intervocalic voicing occurs. The target output [ba.b-a] thus can surface from two different sources: (i) selecting the plural allomorph /bab/ in the input of Stem-A (pl.) (i.e. /bab+a/) and deriving the target faithfully with a final devoicing grammar, and (ii) deriving [ba.b-a] from /bap+a/ via intervocalic voicing as stated above. The current PSI model is thus responsible for monitoring grammar changes along with the variation in SPs to explain the patterns produced at successive learning stages and how a target grammar is (not) acquired.

PSI is in theory compatible with any phonological grammar that allows competitions among different grammar assumptions. In this dissertation, I will adapt a constraint grammar to model different grammar learning stages for its success in modeling child phonology since its invention in Prince & Smolensky (1993/2004) and computational implementation in Tesar (1995). More specifically, an Optimality-Theoretic Gradual Learning Algorithm (OT-GLA; Boersma 1998, Boersma & Hayes 2001) will be implemented to generate surface variations in the acquisition of morphophonology. A similar model would be a GLA learner based on Harmonic Grammar (HG-GLA; e.g. Jesney & Tessier 2011) whose learning outcome may be influenced by gang effects (i.e. more violations of lower-weighed constraints might be worse than a single violation of a higher-weighed constraint). OT-GLA is applied from Ch. 2 to Ch. 5 to develop a computationally simpler prototype of the proposed rich lexicon model along with the input selection process. HG-GLA is discussed in Ch. 6 as a necessary future expansion to incorporate the entire lexical network in a rich lexicon.

With a rich lexicon, it is also possible to develop a learning model based on Exemplar Theory in which grammar is a lexical generalization *per se* rather than a set of additional rules or constraints. Nevertheless, although the idea of constraints has never been formalized in any exemplar model, there are similar concepts; a lexical pressure that requires a surface form to be identical to the exemplars of an exemplar cloud is similar to faithfulness constraints (Kirchner 1999, Kirchner et. al 2010), and articulatory biases (e.g. Wedel 2006) are akin to phonetically-driven markedness constraints. A constraint model with a rich lexicon thus does not highly deviate from an exemplar model and is in fact consistent with a 'hybrid' model proposed in Guy (2007, 2014) and Pierrehumbert (2006) which requires both a rich lexicon and a variable phonological grammar (see §6 of Ch. 1), and the latter is assumed as a stochastic constraint grammar in this thesis.

The following three subsections will elaborate the details of OT-GLA for the upcoming computational simulations.

2.1 Innate and learnable constraints

In Standard OT, a widely accepted assumption is that the constraint set CON is only composed of a fixed number of universal innate constraints. A primary advantage of making such an assumption is to predict possible phonological grammars with different constraint rankings generated from this closed constraint set (i.e. ranking typology), which later motivates Prince (2007) to specifically argues for the typological evidence for including a constraint in an OT grammar. A less stringent criterion for the inclusion of constraints is to have a motivation grounded on articulatory or perceptual complexity (e.g. Hayes et al. 2004; cf. Archangeli & Pulleyblank 1994, Stampe 1979). In sum, different criteria have been made to verify a violable constraint in CON, but the proposed constraints have been commonly assumed as innate.

Some authors, however, propose non-universal constraints created by learners to account for lexical trends and exceptions. For example, Pater (2000, 2009) proposes to deal with lexical exceptions to regular phonological alternations by cloning faithfulness and markedness constraints as higher-ranked constraints lexically-indexed to specific lexical entries to preserve underlying structure or force phonological alternations in their output representation. Becker (2009) adapts this approach and proposes language-specific rankings based on a lexical trend, which can be probabilistically applied to the production of novel forms. For example, in Turkish, the process that changes a stem-final /-t/ to a stem-final [-d] intervocalically only applies to eighteen out of 120 lexical items. In Standard OT, the majority of a stem-final [-t] can be accounted for with the ranking IDENT(voi)-IO » *VtV (i.e. preserving underlying [voi] specifications » no intervocalic /t/; the symbol '»' denotes 'more dominant' or 'outrank'). However, the alternation does not occur in another 102 items with stem-final [-t], the lower-ranked markedness constraint *VtV is only indexed to these 102 items as $*VtV_{102items}$. The markedness constraint is cloned as another markedness constraint $*VtV_{18items}$ that is only indexed to the eighteen items in which intervocalic voicing applies and thus outranks IDENT(voi)-IO. When a novel form with a stem-final [-t] is produced in an intervocalic position, it has a 15% chance (18/120) of violating the top-ranked indexed constraint in the ranking $*VtV_{18items}$ » IDENT(voi)-IO » $*VtV_{102items}$ but a 85% chance (102/120) of violating $*VtV_{102items}$. The probability of the stem-final [-t] variation in novel forms can thus be predicted, and the experimental results in Becker (2009) support this prediction.

Child-specific phonology also suggests non-innate constraints to exist and impose specific restrictions to ease the burden on an immature articulatory system. In Byun (2011), for example, stops and fricatives are found neutralized in a prosodically strong position in child phonology (i.e. onset), contrary to adult phonology, in which phonemic contrasts commonly neutralize in a weak position like coda (e.g. Beckman 1998). Byun suggests that a considerable gesture overlap between onset and nucleus requires independent movements of tongue and jaw to produce a fricative-vowel sequence, which are too demanding for children's articulators. A top-ranked constraint MOVE-AS-UNIT (i.e. independent articulator movements are prohibited) is thus created to relieve the required control for individual articulators. Becker & Tessier (2011) also proposed that child-specific harmonies are results of inducing markedness constraints "in response to the child's own productions and increased articulatory demands" (2011:182).[18]

To a greater extreme, some constraint-based learners simply induce all constraints from learning input without postulating any a priori knowledge of innate constraints such as Alderete et al. (2013), Hayes & Wilson (2008), and van Oostendorp (2014). For example, Hayes & White (2013) followed Hayes & Wilson's (2008) constraint induction learners to develop a constraint grammar of the well-formedness judgments on English-like words (e.g. canift, sneck). Without a restriction that constraints must be innate, unnatural constraints can be added to the grammar freely to prohibit strings that are underrepresented in the lexicon. For example, *[+cont, -strid][-son] (i.e. a cluster of a non-strident fricative followed by a non-sonorant is prohibited) is acquired for less acceptable forms like 'he**thk**er' and 'mu**thp**y'. Hayes & White report that although forms violating unnatural constraints are not as unacceptable as those violating natural ones, there is a trend toward a higher-rating for the forms that conform to unnatural constraints.[19] Recent studies on artificial phonology learning also open the possibility for learnable unnatural constraints. Moreton & Pater (2012a:686) claim that "[…] a learner with substantive bias would acquire phonetically-motivated patterns better than phonetically-arbitrary ones […]", and the term 'better' usually means a bias that allows the former to be learned faster than the latter, rather than a strong claim that the latter is not learnable (see Hayes & White 2013 and Moreton & Pater 2012b for an intensive review of the topic). Thus, to accommodate this growing body of evidence supporting learnable unnatural constraints, I will presume the participation of unnatural constraints in morphophonological acquisition, which may or may not be top-ranked at the end learning state.

This dissertation focuses on a different source of learnable constraints which learners may induce by (re-)analyzing morphophonemic alternations. For example, the derivation /bat+a/→[ba.t-a] succeeds with the target ranking IDENT(voi)-IO » *V[-voi]V (i.e. preserving underlying [voi] specifications » no intervocalic voiceless

---

[18] For more discussion about child-specific phonology, see Inkelas & Rose (2007), Pater (1997), and among many others reviewed in the literature summarized in this section.

[19] Myers & Tsay (2013) provides an alternative analysis of Hayes & White's rating data using two variables Constraint Naturalness and Lexical Typicality (i.e. constraint weight; higher = less typical). The result shows that the difference in rating between violated and non-violated forms only increases with the highly weighed natural constraints (i.e. violating highly weighed natural constraints results in a lower rating).

segment), or the inclusion of a top-ranked unnatural constraint *V[+voi]V (i.e. intervocalic voiced segments are prohibited). Here I take a conservative position that an unnatural alternation must be highly productive to be learned as an unnatural constraint, following the observation that learners do not generalize unnatural alternations as regular patterns easily. I propose the threshold of recognizing unnatural alternations as unnatural constraints to be determined by a function that evaluates the productivity of a specific pattern, dubbed as the Tolerance Principle in Yang (2005) with the theorem (24).

(24) Threshold in the Tolerance Principle (Yang 2005:282), where M_c is the exception threshold for a structural change to be recognized as a regular morphophonological rule, and N is a set of word types with the same structural description, to which the rule can apply.

$$M_c \approx N/\log N$$

For example, assume a structural change 'adding [-d] to all past tense verb stems' as in English and 100 verb stem types (N = 100). The number of the exceptions to the structural change must be lower than 100 / log(100) ≈ 22 (M_c) for the structural change to be recognized as a productive and regular morphophonological rule. Otherwise, the word types with the structural change are assumed to be stored in the lexicon without any rule-application.

Yang (2005:§5.1) claims that the Tolerance Principle can predict the U-shape development of the regular past tense rule in English. In an early acquisition stage, irregular past tense verbs occur more frequently and are thus likely to be picked up by learners first. Assuming that the type frequency of irregular past tense verbs is five and the number of regular ones is two, the regular 'add [-d]' past tense rule cannot be developed as the number of irregular past tense verbs (i.e. five) exceeds the exception threshold predicted by Tolerance Principle (i.e. M_c = (2 + 5) / log(2 + 5) = 3.6). In this case, regular past tense verbs are simply memorized as if they are not related, and produced correctly. Later, assume that the lexicon expands with more regular past tense verbs (e.g. ten), but the number of exceptions remains unchanged (i.e. five). The threshold is now raised to exceed the number of exceptions to the past tense rule (M_c = (10 + 5) / log(10 + 5) = 5.5), which suggests the past tense rule to be sufficiently productive, and the rule is applied across the board by default. This stage thus features the overgeneralization patterns in which the past tense rule is extended to irregular verbs by learners. At the final stage, learners learn to apply this past tense rule productively but at the same time make a list of exceptions to this rule (i.e. irregular past tense verbs). The overgeneralization patterns thus disappear. This prediction is borne out with the child language data in Marcus et al. (1992).

In PSI-OT-GLA, an 'exception' to an unnatural constraint like *V[+voi]V can be defined as every correct output type violating the constraint. The algorithm keeps updating the type frequency of correct outputs with an intervocalic obstruent (i.e. types; N) and the type frequency of correct outputs with a voiced intervocalic obstruent (i.e. exception; M_c); whether intervocalic devoicing is productive enough to be induced as a constraint can thus be decided at any given point. For example, assuming that in a learning cycle there are 100 correct output types with an intervocalic obstruent, the obstruent must be voiceless at least in 78 of the 100 output types for the unnatural constraint *V[+voi]V to be included in the constraint set.

PSI-OT-GLA is not programmed with any built-in constraint inducer that can freely generate unnatural constraints following Tolerance Principle in a simulation. Instead, the algorithm simply includes a small number of relevant unnatural constraints by default in our case studies of morphophonological acquisition, which are assumed possible in a reasonable constraint induction space. For example, Hayes & Wilson (2008) claims that segmental constraints may have at most three natural class feature matrices for a smaller constraint induction space, and the unnatural constraint *V[+voi]V follows this criterion if it is represented with feature matrices (i.e. *[+syllabic][-sonorant, +voice][+syllabic]. In PSI-OT-GLA, therefore, an unnatural constraint is simply 'turned on' or 'turned off' depending on the number of output types that violate the constraint; when an unnatural constraint is 'turned off', its violation marks are not evaluated by the algorithm. It is our hope to equip PSI-OT-GLA with a constraint inducer in near future to further approximate real learners' behaviour, but since this part of the algorithm is not of primary interest in this thesis, it is omitted for now.

2.2 Initial state of constraints and re-ranking process

With a set of constraints, either innate or unnatural, the next issue is the initial state of constraint ranking from which learners start to approach the target ranking, and a number of studies suggest that constraints are not ranked equally in the initial state of a constraint grammar. Menn (1980:35-36, see also Stampe 1979) claims that learning the target phonological grammar is a gradual relaxation of output constraints, since children usually produce unmarked and structurally simpler forms in their early development of phonology. In OT, it has been commonly assumed that markedness constraints, which are only violated by output candidates, initially outrank Input-Output (IO) faithfulness constraints (e.g. Smolensky 1996, Davidson et. al 2004; cf. Hale & Reiss 1998).[20] The term 'relaxation' in OT can be expressed by demoting some initially higher-ranked markedness constraints below IO faithfulness constraints to allow the production of more marked forms. As shown in Tableau 2.1, the initially higher-ranked markedness constraint *A prohibit any output containing the phonological element A, thus forbidding the input /A/ to be produced faithfully as the more marked correct output (denoted by '✓'). Rather, an output error (denoted by '☹') surfaces despite the deviation from the input causing the violation of the faithfulness constraint IDENT-IO. For the correct output to be produced, constraints in favor of the correct output, *A in this case, are demoted (denoted by '→'), whereas constraints violated by output errors, here IDENT-IO, are promoted (denoted by '←'). Consequently, the reverse ranking allows the correct output with the more marked element to surface as illustrated in Tableau 2.2.

/A/	*A	IDENT-IO
✓A	*!→	
☞☹B		←*

Tableau 2.1. Demotion of markedness constraint and promotion of faithfulness constraint

/A/	IDENT-IO	*A
☞✓A		*
B	*!	

Tableau 2.2. Production of correct output after constraint re-ranking

[20] Hayes (2004) and Tessier (2007, et. seq) also propose that Output-Output (OO) faithfulness constraints (Benua 1995, 1997) initially outrank IO faithfulness constraints and should at least tie with markedness constraints to explain paradigmatic uniformity in early child morphophonology. OO faithfulness constraints will be redundant in the following simulations but readers can refer to Ch. 6 for a discussion of extending OO faithfulness constraints in a rich lexicon model.

In PSI-OT-GLA, each constraint has a constraint value and a constraint generally outranks another constraint by having a higher constraint value. To capture the initial Markedness » IO-Faithfulness ranking, the constraint value of markedness constraints is set to 100 and that of IO faithfulness constraints is set to zero. Unnatural markedness constraints that are included in the algorithm by default, if 'turned on' following Tolerance Principle (see §2.1 above), will be initially ranked at the top (i.e. equal to the highest constraint value in the grammar; see also Becker & Tessier 2011).

2.3 Gradual changes and random noise in constraint values

In PSI-OT-GLA, adjustments are made to constraint values only when an output error is produced, which is the well-known error-driven learning procedure (e.g. Wexler & Culicover 1980) adapted in various constraint-based learning algorithms. Gradual constraint promotion and demotion are manipulated by adding or subtracting a small amount from constraint values, which is called plasticity. In the original OT-GLA framework (e.g. Boersma 1998), plasticity is fixed for a symmetrical constraint promotion and demotion (i.e. constraint values are raised or lowered by the same amount). To ensure restrictive learning and avoid the Subset Problem (Angluin 1980 and Baker 1979), however, an asymmetry between promotion and demotion is required. Jesney & Tessier (2011) propose an asymmetry that the plasticity of IO faithfulness constraints should be smaller than that of markedness constraints, and Magri (2012) suggests that the plasticity should be smaller for promoted constraints than for demoted constraints regardless of constraint type.

The current PSI model will adapt Magri's asymmetry formula (25) since it is designed specifically for restrictive learning in OT.[21] All in all, the plasticity for constraint demotion will be fixed at 0.1 (as in Boersma 1998 and Boersma & Hayes 2001), and the promotion plasticity is smaller than 0.1, depending on the number of demoted undominant loser-preferrer (i.e. undominant constraints violated by the winner but not the loser) and promoted winner-preferrer (i.e. constraints violated by the loser but not the winner) in each constraint promotion and demotion process.

[21] Magri's (2012) demotion bias was originally implemented in a non-probabilistic version of GLA since random noises added to raw constraint values were ignored. Nevertheless, Magri explained that the convergence results with this demotion bias can be easily extended to a probabilistic ranking system (2012:247, fn. 9).

Since constraint violations might be inconsistent in different learning cycles, the promotion plasticity may thus vary from time to time.[22]

(25) Asymmetry between promotion and demotion plasticity based on Magri (2012:217)

$$promotion\ amount < 0.1 \times \frac{number\ of\ demoted\ undominant\ constraints}{1 + number\ of\ promoted\ constraints}$$

$$demotion\ amount = 0.1$$

Another variable that can cause a shift in constraint values is evaluation noise, which is a random number added to each constraint value in each evaluation process of the optimal output. The shift is not permanent; a random noise is added to each constraint value to probabilistically determine the ranking between constraints in every learning cycle. The closer are two constraint values, the more likely is the ranking between the two corresponding constraints are changed by a random noise. The current PSI-OT-GLA algorithm generates a noise value for each constraint from a Gaussian distribution with a mean of zero and a standard deviation of two.

Three important issues in the constraint re-ranking process need to be addressed as well. First, since the ranking between some constraints are intrinsically fixed for phonological implications, the gradual constraint promotion and demotion process should not change the ranking of these constraints. The original framework of OT (i.e. Prince & Smolensky 1993/2004) already includes such rankings to account for syllabification in Berber with sonority scales as in (26).

(26) A sonority-based intrinsic ranking for syllabification

*PEAK/OBS » *PEAK/FRICTIVE » *PEAK/NASAL » *PEAK/LIQUID » *PEAK/GLIDE » *PEAK/HIGH » *PEAK/MID » *PEAK/LOW

(i.e. lower sonority sounds are less preferred as the nucleus of a syllable and thus violates intrinsically higher-ranked *PEAK constraints.)

[22] Boersma (1998:274) also proposes to gradually reduce plasticity to model a stabilized constraint ranking of a matured speaker. Nevertheless, to my best knowledge, no conclusive findings support the grammar maturity over the learning span of morphophonology, and thus the plasticity remains unchanged in the following series of simulations. In addition, it might also be plausible to assume that the constraint ranking remain as flexible as possible in morphophonological acquisition to allow learners to change and test different grammar assumptions.

Other intrinsic rankings may be grounded on phonetically driven factors such as articulatory and perceptual complexity (e.g. Hayes et. al 2004) or morphological hierarchy (e.g. stem faithfulness is more important than affix faithfulness: FAITH-IO_{ROOT} » FAITH-IO_{AFFIX} in McCarthy & Prince 1995). While the discussion of the role of intrinsic rankings in the GLA is not the primary focus in this dissertation, the fixed ranking will be required to produce a correct learning result of the diachronic tonal change in Mandarin in Ch. 4. To simulate a fixed ranking between two constraints in the GLA, every two intrinsically-ranked constraint values are separate with a fixed interval. For example, if the lower-ranked constraint in an intrinsic ranking is slightly promoted, the higher-ranked constraint is automatically promoted by the same amount, so the difference between the two constraint values remains constant.

Second, previous OT work rarely discusses whether constraint values should have a roof and a floor. In this dissertation I propose that morphophonological acquisition with the demotion bias in (25) can only be successful without a floor (i.e. the lowest constraint value) in the GLA.[23] Therefore, the initial constraint value of IO faithfulness constraints is zero does not mean that the lowest possible constraint value is also zero; i.e. constraint values can be negative. The necessity of this implementation is illustrated with the following example.

When the allomorph /bap/ of Stem-A is selected as the input in the plural context, requires the ranking *V[-voi]V » IDENT(voi)-IO is required to derive the correct output (i.e./bap+a/→[ba.b-a]). However, to faithfully derive the correct output of 'Stem-B (pl.)' (i.e. /bat+a/→[ba.t-a]), the opposite ranking IDENT(voi)-IO » *V[-voi]V is required; both constraints thus fall in cycles of demotion and promotion. Ideally, learners should identify /bap/ as a worse input option since it largely surfaces as output errors when IDENT(voi)-IO outranks *V[-voi]V. Without selecting /bap/ in the plural context, there is no pressure to promote *V[-voi]V over IDENT(voi)-IO; learners thus should be able to fully rebuild the target ranking IDENT(voi)-IO » *V[-voi]V at some given point.

However, the primary concern is that with a demotion bias, the target ranking can only be rebuilt by demoting *V[-voi]V further, and this might be impossible if

[23] Nevertheless, as pointed out by Magri & Storme (2013), such a demotion bias without a floor may fail to capture simple variations which can be modeled by an unbiased GLA in Boersma & Hayes (2001). The balance between restrictive learning and variation modeling thus remains an issue to be further investigated.

there is a floor for a constraint value. Assume that both constraint values start at fifty and that both constraint values are promoted and demoted for twenty times. If plasticity is equal for promotion and demotion, the two constraint values will remain at fifty. However, with a demotion bias like +0.1 for promotion and -0.2 for demotion, the two constraint values are lowered to 50 – (0.1 – 0.2) × 20 = 40. If the competition between the two constraint rankings lasts, both constraint values may drop to the floor, say zero, at which point grammar learning stalls before the better input option is identified. The removal of the floor solves this problem since the two constraint values can be demoted below zero, and the algorithm can always have the chance to re-established the target ranking IDENT(voi)-IO » *V[-voi]V by eventually demoting *V[-voi]V more significantly than IDENT(voi)-IO.

Finally, recall that since PSI learners are not equipped with a preliminary knowledge of the correct inputs and grammar, the learners do not know whether the selection of inputs or the grammar should be responsible for producing output errors. The learners thus make adjustments to both lexical variables (IEP and MEP) and the current constraint ranking in the error-driven learning procedure. For instance, when *[ba.p-a] 'Stem-A (pl.)' is derived from the 'incorrect' input /bap+a/ with the constraint ranking IDENT(voi)-IO » *V[-voi]V, the constraint value of IDENT(voi)-IO and *V[-voi]V are pulled closer for the opposite ranking *V[-voi]V » IDENT(voi)-IO as mentioned above, and $IEP_{/bap/}$ is also raised to lower $SP_{/bap/}$. As the above process repeats itself, there are two possible end states. On the one hand, $SP_{/bap/}$ may be lowered before the opposite ranking *V[-voi]V » IDENT(voi)-IO is stabilized. On the other hand, the reverse ranking is stabilized as a diachronic shift triggered by selecting a different basic paradigm. Both ends will be captured in simulations based on PSI-OT-GLA from Ch. 3 to Ch. 5. Indeed, the ideal solution is to recognize the input as the troublemaker and not to adjust the target constraint ranking, but such a solution is less feasible due to tremendous computational efforts (see §6.1), and besides the 'correct' input cannot always be perfectly recognized as reviewed in Ch. 1.

2.4 Gradual Learning Algorithm: Caveats

Adopting the GLA as the grammar learning module does not mean ignoring its intrinsic problems discovered previously. For example, Pater (2008) found that in the original version (i.e. Boersma 1998 and Boersma & Hayes 2001), the GLA fails to converge on a specific grammar and constraint values grow infinitely. This potential problem is solved following the adoption of Margi's (2012) demotion bias in PSI-OT-GLA.

Another algorithmic issue in the GLA is related to its nature of constraint demotion and promotion directly controlled by the frequency of violating the constraint (Tessier 2009): If a constraint is violated more frequently by an output error than another constraint, the former is promoted more frequently than the latter. A general IO faithfulness constraint (e.g. MAX-IO = No deletion of input elements) is thus promoted faster than a positional IO faithfulness constraint (e.g. MAX-σ́ = No deletion of input elements in stressed syllables), since the latter is only violated in a specific context. An intermediate stage preserving underlying elements in specific positions with the ranking Pos-IO-Faithfulness » Markness » IO-Faithfulness is thus impossible in the GLA. This dissertation does not seek to deal with these problems but considers the GLA a simple application of an on-line learning system that allows us to easily monitor the grammar development in morphophonological acquisition. Needless to say, it is hoped to ultimately reconcile the proposed PSI-OT-GLA with the solution to the problems in the GLA in the future.

3. Two-stage (morpho-)phonological acquisition

So far, I have focused on the interaction between the input selection process and its interaction with grammar learning, but it is important to note that such an interaction does not occur at an early stage in language acquisition due to the relatively late emergence of morphophonological knowledge. Various studies found that phonotactic knowledge develops rapidly in learners' early infancy; when learners are as young as 9 months, they have already extracted and adapted the major phonotactic patterns in their learning inputs. The experimental results in Jusczyk et al. (1993) lead to the conclusion that nine-month-old English-learning infants are able to distinguish English stress patterns from Dutch ones. English learning infants of the same age also show their preference for initially stressed syllables, which are the major prosodic patterns in English (Jusczyk, Cutler, & Redanz 1993). On the other hand, if phonetic differences are not phonologically contrastive in their learning inputs, the perceptual sensitivity to these differences reduces quickly. Spanish learning infants can distinguish word-final stress from word-initial stress when they are nine-month-old, whereas French learning infants of the same age are naïve to the stress difference since stress is completely predictable in French (Skoruppa et al. 2009). English learning children fail to perceive a three-way VOT contrast in Hindi after the age of four whereas six-month-old English learning infants are highly sensitive to the contrast (Werker & Tees 1984).

Compared to the early development of their phonotactic knowledge, language learners do not seem to have any morphological knowledge until at least fifteen months old, which is the age with the correct use of the accusative suffix [-a]~[-e] by Turkish learning children observed in Aksu-Koc & Slobin (1985). More case studies on morphophonological learning show an age threshold much higher than that of mastering the Turkish suffixes. Modern Hebrew (Berman 1985) and English learning children cannot produce morphologically complex forms correctly until at least four years of age (Berko 1958). Other experimental studies also document a further delay due to a low type frequency of a specific morphophonemic pattern. In Hungarian, vowel lengthening occurs in stem-final vowels [a] and [e] when there is a suffix attached to the stem, but the stems with final [e] are less frequent and stem-final [e] lengthening is not even fully acquired by seven-year-olds (MacWhinney 1978). A similar delay was also found in the elicitation tasks in Kerkhoff (2007) and Zamuner et al. (2012); Dutch-learning children still produce a certain amount of plural form devoicing errors like *[ba.p-a] (cf. [ba.b-a]) in the hypothetical language at the age of seven. In sum, the threshold may be relevant to various aspects of target morphophonological patterns and is open to future study,[24] but the chronological order of the two distinct stages should be rather uncontroversial.[25]

One major difference in the two learning stages in PSI-OT-GLA is that the input selection process is assumed not active in the phonotactic learning stage (P stage) to interact with grammar learning, given that learners' cognitive function has not fully been developed to associate morphologically related forms and decompose morphologically complex forms as individual morphemes (see also Hayes 2004;

---

[24] Baer-Henney & van de Vijver (2012) discuss the complexity of morphophonemic alternations defined by substance, locality, and amount of exposure of morphophonemic alternations, which may contribute to different time lengths required for acquiring morphophonological patterns.

[25] It is worth noting that phonotactic studies mainly focus on perception tasks whereas morphophonological studies interpret results of production tasks. Since production is usually more delayed than perception and comprehension, there is a potential confound that the delay observed in morphophonological studies stems from the delay in production, rather than the delay in learning morphophonemic alternations. Studies such as Zamuner et al. (2006) nevertheless found that children are less capable of relating forms with morphophonemic alternation together, which suggest that the delay in morphophonological acquisition roots in comprehension.

Tesar & Prince 2003; Jarosz 2006b for similar models that exclude lexical learning from the P stage). The input of any stored morphologically complex forms in the P stage, following Lexicon Optimalization in OT, is always identical to the output; the input of [ba.b-a] 'Stem-A (pl.)' is thus always /baba/ and the input of [bap] 'Stem-A (sg.)' is always /bap/. In terms of lexical variables, SP, IEP, and MEP will not be calculated in the P stage either. Furthermore, without morphological decomposition, it is assumed that learners can only track the token frequency of the whole perceived input but do not relate this frequency information to any specific surface allomorph of a morpheme. That is to say, in the M stage, it is assumed that the token frequency of an allomorph begins with zero. Learners are assumed to have the ability to induce unnatural markedness constraints in the P stage, but due to the lack of morphological knowledge, these constraints must be free of morphological structures (e.g. morpheme boundary; see Ch. 5).

Phonotactic learning is not undermined without morphological decomposition if the learner is biased to acquire a most restrictive constraint grammar (see §2.2 and §2.3). If we only consider the four surface forms [ba.b-a], [ba.t-a], [bap], and [bat] in the hypothetical language, the absence of any voiced coda can be immediately observed from this subset. With the initial Markedness » IO-Faithfulness ranking bias, *VOICEDOBSCODA should always outrank IDENT(voi)-IO and will not be demoted through the learning process. On the contrary, while initially higher-ranked, *V[-voi]V will be gradually demoted below IDENT(voi)-IO in the constraint hierarchy for the positive evidence of intervocalic voiceless segments (e.g. [ba.t-a]). The target ranking *VOICEDCODA » IDENT(voi)-IO » *V[-voi]V can thus be acquired without being sensitive to any morphological structure.

The constraint ranking acquired in the P stage will be the initial constraint ranking after morphological awareness (M stage), which is still subject to the adjustments possibly triggered by the input selection process mentioned earlier; selecting /bap+a/ as the input of 'Stem-A (pl.)' requires *V[-voi]V to outrank IDENT(voi)-IO to derive the target [ba.b-a], and the two constraint values are pulled closer. Nevertheless, the advantage of the inherited constraint ranking from the previous stage is to allow allomorphs which can produce targets with the acquired grammar to gain a higher SP.[26] That is, with the initial ranking IDENT(voi)-IO » *V[-

[26] See also Tesar & Prince (2003) for how the ranking acquired in the P stage helps achieve the goal of probing UR in non-stochastic OT.

voi]V, [ba.b-a] can only be derived from /bab+a/ but not /bap+a/, and $\text{IEP}_{/bab/}$ is thus lowered for a higher $\text{SP}_{/bab/}$.

4. Summary of PSI-OT-GLA learning process at different stages

This section serves as a summary of the above learning process in PSI-OT-GLA by walking through the learning process step by step in the two distinct learning stages respectively in Figure 2.2 and Figure 2.3.

At the beginning of each learning cycle in the P stage simulation in Figure 2.2, learners randomly perceive one token from input source (e.g. adult speaker) based on the chance for the source to produce the token, which is [ba.p-a] in Cycle 1. Learners then update the frequency of /bapa/ in their lexicon. In the next step, learners randomly select one lexical item to produce. Since /bapa/ is the only lexical item with a non-zero token frequency, learners necessarily choose to produce this form. Assuming that learners produce the output error *[baba] with intervocalic voicing due to the initial markedness » faithfulness ranking bias, learners have to promote IDENT(voi)-IO and demote *V[-voi]V. At the end of each cycle, each stored token slightly decays from the memory. Proceeding to Cycle 2, learners again perceive another token from the input source, this time [ba.b-a], and update the frequency of /baba/. With the same token frequency, /bapa/ and /baba/ have the same probability to be produced by learners. Hopefully, learners can produce either form as a correct output so no grammar change needs to be made before moving to the next learning cycle.

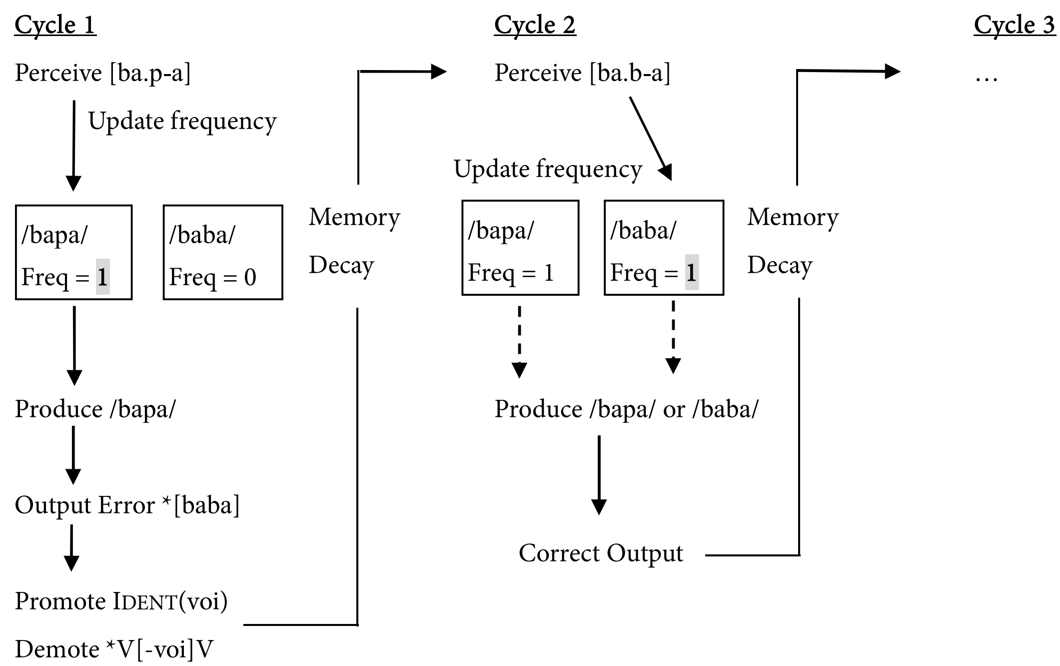


Figure 2.2. Learning cycles in the P stage in PSI-OT-GLA; shaded numbers represent updated lexical information

The M stage simulation in Figure 2.3 starts with perceiving learning inputs randomly as in the P stage simulation, and in Cycle 1 [bap] 'Stem-A (sg.)' is assumed as the learning input. Following this, learners update the token frequency of the lexical item 'Stem-A (sg.)' (as well as the singular allomorph /bap/). If no other forms are perceived by learners at this point, learners must choose to produce 'Stem-A (sg.)'. Input selection then determines the input of this form, which can only be /bap/ due to the lack of input competitors. Note IEPs and MEPs are initially 0.5 with the smoothing process (0.5/1 = 0.5). When learners successfully produce the correct output, they immediately lower $IEP_{/bap/}$ and $MEP[X+\emptyset]_{SG}$ to 0.25 (i.e. 0.5 / (1 + 1) = 0.25).

In Cycle 2, learners now perceive a token of [ba.b-a] 'Stem-A (pl.)' and update the token frequency accordingly. With an equal token frequency in the lexicon, learners may produce either the singular or plural form in Cycle 2, and let's assume 'Stem-A (pl.)' to be the choice. Since the plural allomorph /bab/ now has a non-zero token frequency, it can compete with the singular allomorph /bap/ as the input of Stem-A. Thanks to a lower $IEP_{/bap/}$ and $MEP[X+\emptyset]_{SG}$, however, the singular allomorph has a dominant SP of 0.8, and learners may prefer /bap+a/ as the input of 'Stem-A (pl.)'. As demonstrated in previous sections, this input choice may lead to the output error *[ba.p-a], and learners will have to do two things: Firstly, they increase the corresponding IEP and MEP (i.e. (1 + 0.5) / (1 + 1 + 1) = 0.5). Secondly, the output error prompt them to promote *V-[-voi]V and demote IDENT(voi)-IO to allow the derivation /bap+a/→[ba.b-a]. In the rest of morphophonological acquisition, learners will continue to update the lexical information and change constraint ranking in order to produce as many correct outputs as possible.

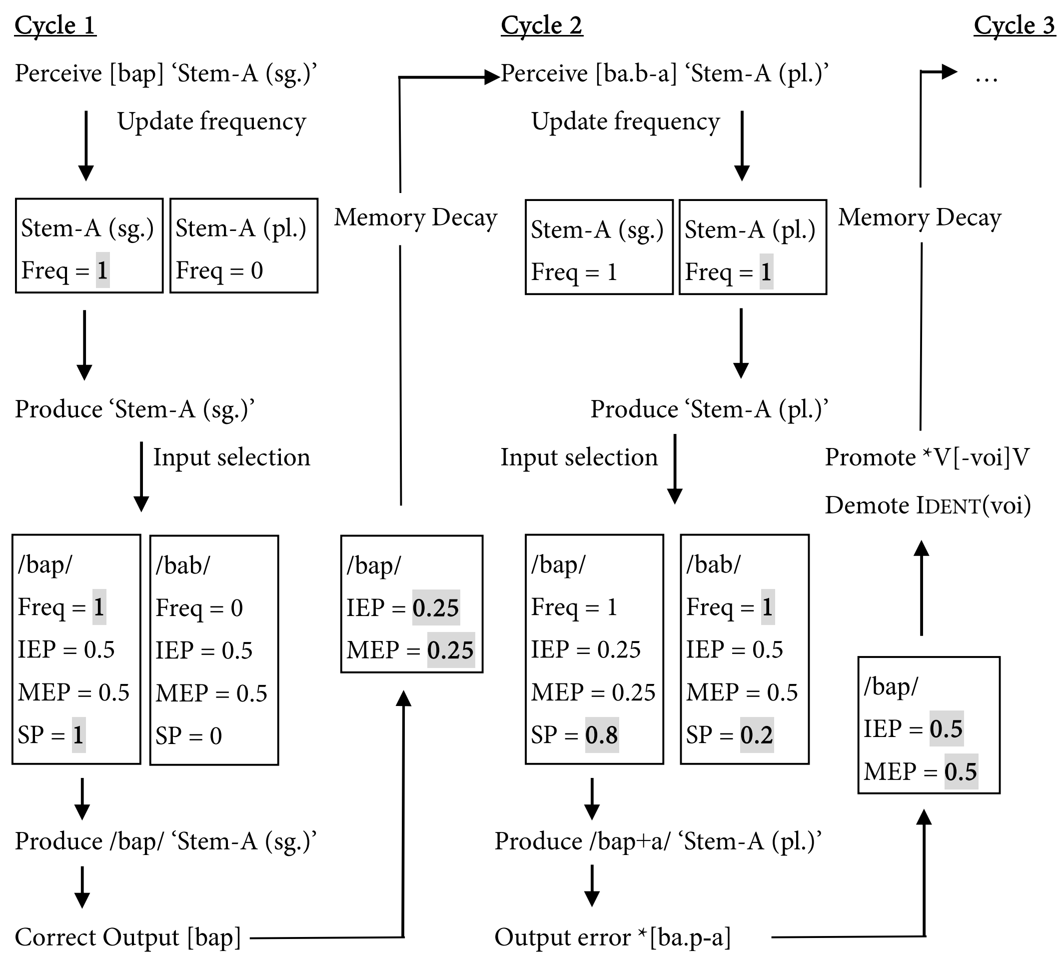


Figure 2.3. Learning cycles in the M stage in PSI-OT-GLA; shaded numbers represent updated lexical information

5. Testing PSI-OT-GLA with toy data

This section includes a series of simulations based on the toy data from the hypothetical language to illustrate the acquisition of the target constraint grammar in the two different learning stages, the interaction between the input selection process and the constraint re-ranking process, and how the SP of plural allomorphs becomes (or fails to become) more dominant with different frequency biases toward singular allomorphs in the input selection process. The simulation was compiled as a Java® program with Eclipse Standard version 4.3.2 (The Eclipse Foundation 2014).[27]

5.1 Training corpus and input distribution

The original data set (8) in the hypothetical language is now expanded in (27) to include two more allomorph pairs /bas/~/baz/ and /bak/~/bag/ to demonstrate the global effect of MEP illustrated in Figure 2.1 in §1.3.

(27) Expanded data set in the hypothetical language

[bap]	'Stem-A (sg.)'	[ba.b-a]	'Stem-A (pl.)'
[bat]	'Stem-B (sg.)'	[ba.t-a]	'Stem-B (pl.)'
[bas]	'Stem-C (sg.)'	[ba.z-a]	'Stem-C (pl.)'
[bak]	'Stem-D (sg.)'	[ba.g-a]	'Stem-D (pl.)'

The input distributions of [bap] and [ba.b-a] with the three frequency biases are shown in Table 2.1. Furthermore, I will assume that each of the forms [bat] and [ba.t-a] occupies 25% of the three input distributions and each of the forms [bas], [ba.z-a], [bak], and [ba.g-a] occupies 8.33% of each distribution. In sum, only the proportion of [bap] and [ba.b-a] varies for the three frequency biases. Memory decay will not be involved in the following toy data simulation to simplify the demonstration of PSI-OT-GLA and the discussion of learning results.

	frequency ratio between [bap] and [ba.b-a]		
	1:1	9:1	19:1
[bap]	8.33%	15.01%	15.85%
[ba.b-a]	8.33%	1.67%	0.83%

Table 2.1. Input distributions with three frequency ratios of [bap]:[ba.b-a]

5.2 Constraint set, initial state, and target grammar

Only eight innate constraints listed in (28) are involved in the following simulations. Markedness constraints initially outrank IO faithfulness constraints with an initial

[27] The source code is available at http://hdl.handle.net/10402/era.39158.

constraint value difference 100 vs. 0. Unnatural markedness constraints are not included in the following toy-data-based simulations for a simple demonstration of sequential changes in lexical variables and constraint grammar in PSI-OT-GLA.

(28) Constraints for grammar learning in the hypothetical language

a. ONSET: Syllables without onset are prohibited.
b. *VOICEDOBS: Voiced obstruents are prohibited.
c. *VOICEDOBSCODA: Voiced obstruent codas are prohibited.
d. *CODA: Syllable codas are prohibited.
e. *V[-voi]V: Intervocalic voiceless consonants are prohibited.
f. MAX-IO: Every input segment must have an output correspondent.
g. DEP-IO: Every output segment must have an input correspondent.
h. IDENT(voi)-IO: Every input specification of [voice] must be preserved in the output.

The target ranking is expected to have ONSET at the top of the ranking hierarchy since none of the eight hypothetical forms is onsetless. IDENT(voi)-IO is expected to be outranked only by MAX-IO and DEP-IO since a voiced coda is repaired neither with deletion nor with epenthesis. IDENT(voi)-IO should be dominated by *VOICEDOBSCODA. Finally, *VOICEDOBS, *CODA, and *V[-voi]V should be ranked at the bottom due to the presence of voiced obstruent onsets, coda consonants, and intervocalic voiced segments. The initial and target ranking are summarized in (29).

(29) Initial and target ranking in toy data simulations

Initial Ranking: {ONSET, *VOICEDOBS, *VOICEDOBSCODA, *CODA, *V[-voi]V} »
{MAX-IO, DEP-IO, IDENT(voi)-IO}

Target Ranking: {ONSET, *VOICEDOBSCODA, DEP-IO, MAX-IO} »
IDENT(voi)-IO » {*VOICEDOBS, *CODA, *V[-voi]V}

5.3 Results: Grammar learning

Each simulation with one of the three distributions in Table 2.1 has two different learning stages as reviewed in §3. A few different numbers of learning cycles were tested, and the learning outcomes below were taken from simulations with 20,000 P stage learning cycles and 50,000 M stage learning cycles. These outcomes are adopted simply because the target phonotactic knowledge can be acquired in all three conditions after 20,000 P stage learning cycles and distinct morphophonological learning outcomes in the three conditions can be generated within 50,000 M stage learning cycles.

The discussion of simulated results below starts with phonotactic learning in the P stage where lexical factors are irrelevant. The constraint values at the end of the P stage are translated as final constraint rankings with strict domination in Table 2.2 for the comparison between the acquired ranking and the target ranking; if the difference between two constraint values is more than five, the ranking between the two constraints is assumed to be strict since the probability for the constraint with a lower constraint value to dominate the one with a higher value ranking is very low.

Bias Ratio	
1:1	Ranking: {ONSET, *VOIOBSCODA}» IDENT(voi) » {*V[-voi]V, *VOIOBS} » {MAX, DEP} » *CODA Value: 100.1 100 57.6 49.4 48.1 26.5 23.3 14.1
9:1	Ranking: {ONSET, *VOIOBSCODA}» IDENT(voi) » {*V[-voi]V, *VOIOBS} » MAX » DEP » *CODA Value: 100.2 100 58.2 49.5 48.7 31.1 21.1 11.8
19:1	Ranking: {ONSET, *VOIOBSCODA}» IDENT(voi) » {*V[-voi]V, *VOIOBS} » MAX » DEP » *CODA Value: 100.3 100 58 49.4 48.6 30.9 21.4 11.8

Table 2.2. Constraint values/rankings acquired with three bias ratios in the P stage simulation

With the three different bias ratios toward [bap], the three acquired constraint rankings at the end of the P stage are essentially identical but slightly different from the target ranking in §5.2. The identical sub-rankings are IDENT(voi)-IO » *V[-voi]V (triggered by the mapping /bata/→*[ba.da]), IDENT(voi)-IO » *VOICEDOBS (triggered by the mappings like /bat/→*[pat] and /baba/→*[pa.pa]), {DEP-IO, MAX-IO} » *CODA (triggered by the mappings like /bat/→*[ba] and /bat/→*[ba.ta]), and top-ranked ONSET and *VOICEDOBSCODA (for the absence of onsetless syllables and voiced obstruent codas).

The crucial ranking difference is IDENT(voi)-IO » {MAX-IO, DEP-IO} in Table 2.2 but {MAX-IO, DEP-IO} » IDENT(voi)-IO in the expected target ranking. The reason for the emergence of this ranking is the lower pressure of promoting MAX-IO and DEP-IO: In theory, MAX-IO should be promoted over *VOICEDOBS to avoid deleting a voiced obstruent onset (e.g. /bat/→*[at]). However, deleting an onset is banned by the top-ranked ONSET and it is unnecessary to further promote MAX-IO to do the job. The promotion of DEP-IO is only required at the very beginning to dominate *CODA to avoid inserting a vowel after a potential coda consonant (e.g. /bat/→*[ba.ta]). Therefore, the two faithful constraints are not promoted as much as IDENT(voi)-IO in the P stage simulation.

This sub-ranking, however, wrongly predicts epenthesis or deletion, rather than devoicing, of a voiced coda in the input with a top-ranked *VOICEDOBSCODA as in Tableau 2.3. This result, however, is not surprising since inputs in the P stage simulation are always identical to their target (i.e. Lexicon Optimization), and voiced obstruent codas are thus absent from the inputs. Learners thus cannot produce a deletion error like *[ba] or an epenthesis error like *[ba.ba] from the input /bab/, which can trigger the promotion of MAX-IO and DEP-IO and the demotion of IDENT(voi)-IO. In sum, although the restrictive learning strategy can keep *VOICEDOBSCODA top-ranked without the presence of voiced obstruent codas, the IO faithfulness constraints cannot be ranked specifically to repair an underlying voiced obstruent coda with devoicing at this point.[28]

/bab/	*VOIOBSCODA	IDENT(voi)	MAX	DEP
☞☹ba			*	
☞☹ba.ba				*
✓bap		*!		
bab	*!			

Tableau 2.3. Coda deletion error with the rankings acquired in the P stage; '✓' = correct output, '☹' = output error

The constraint values at the 10,000th cycle in the M stage simulation are also translated as constraint ranking in Table 2.3. Regardless of the bias ratio, the PSI-OT-GLA acquires the target ranking in which MAX-IO and DEP-IO outranks IDENT(voi)-IO. As explained above, the re-ranking between the three constraints does not occur until an input with a voiced obstruent coda like /bab/ is available in

[28] Recall that the speed of promoting and demoting a constraint in the GLA is determined by the frequency of violating the constraint by output errors and targets (see §2.4). Therefore, if the learning inputs include more items that produce output errors violating MAX-IO and DEP-IO, any of the two constraints might be promoted over IDENT(voi)-IO in phonotactic learning. Due to this frequency-sensitive factor, I will not conclude that IDENT(voi)-IO » {MAX-IO, DEP-IO} must be the unique constraint ranking at this point and that the errors *[ba] and *[ba.ba] must both emerge in the M stage when an input option like /bab/ is available.

the production of singular forms and competes with /bap/ via the input selection process. The mappings /bab/→*[ba] and /bab/→*[ba.ba] then trigger the promotion of MAX-IO and DEP-IO respectively and the demotion of IDENT(voi)-IO.

Bias Ratio	
1:1	Ranking: {ONSET, *VOIOBSCODA}» MAX » DEP » *CODA » {IDENT(voi), *V[-voi]V*, VOIOBS} Value: 100.1 100 35.7 28.2 14 10.5 8.5 4.9
9:1	Ranking: {ONSET, *VOIOBSCODA}» MAX » DEP » {IDENT(voi), *CODA, *V[-voi]V*, VOIOBS} Value: 100.2 100 36.4 24.5 12.4 11.8 10.5 7.8
19:1	Ranking: {ONSET, *VOIOBSCODA}» MAX » DEP » {IDENT(voi), *V[-voi]V*, *CODA, VOIOBS} Value: 100.3 100 35.5 24.6 14 12 11.8 9

Table 2.3. Constraint values/rankings acquired at the $10{,}000^{th}$ cycle in the M stage simulation

Note that at this stage, IDENT(voi)-IO has not fully dominated *V[-voi]V since at times *V[-voi]V must outrank IDENT(voi)-IO to derive [ba.b-a] from the input /bap+a/. In theory, IDENT(voi)-IO should gradually dominate *V[-voi]V after $SP_{/bap/}$ is lowered and the number of selecting inputs like /bap+a/ decreases. In below, we will see how plural allomorphs receive a dominant SP along with successful grammar learning in the M stage simulation, and how the algorithm could fail to converge on the 'correct' allomorph and constraint ranking with a stronger bias toward singular allomorphs.

5.4 Results: Lexical learning

The discussion of the lexical learning results in the M stage simulation will focus on $SP_{/bab/}$ (selection probability of /bab/), $IEP_{/bab/}$, (Individual Error proportion of /bab/) and $MEP[X+a]_{PL}$ (Morphological Error Proportion of the context $[X+a]_{PL}$) since the allomorph /bab/ is the target to be assigned a high SP. I will first start with the learning context where there is no frequency bias toward singular allomorphs (i.e. 1:1).

In Figure 2.4, it is shown that $SP_{/bab/}$ changes significantly at the incipient stage since final devoicing is not acquired in the P stage simulation as discussed in the previous sections. Therefore, while the input /bab+a/ can surface faithfully as the correct output [ba.b-a], which raises $SP_{/bab/}$, output errors like *[ba] or *[ba.ba] wrongly emerged from /bab/ via deletion or vowel epenthesis lower $SP_{/bab/}$. The turning point of SP, IEP, and MEP occurs right after the $5{,}000^{th}$ learning cycle. The reason for $IEP_{/bab/}$ and MEP[X+a]$_{PL}$ to drop after this point is the dominance of MAX-IO and DEP-IO; when the two constraints outrank IDENT(voi)-IO, deletion and epenthesis errors of singular forms like *[ba] and *[ba.ba] stop emerging from /bab/, and the input hereafter generates no more output errors. On the contrary, the allomorph /bap/, if selected as the stem input of the plural form (i.e. /bap+a/), the output error *[ba.p-a] is derived with the ranking IDENT(voi)-IO » *V[-voi]V. These developments thus altogether generate an $SP_{/bab/}$ of almost 1. Since the constraint ranking is generated stochastically, same output errors may surface sporadically, which at times lower $SP_{/bab/}$ as indicated by the vertical blips in the blue curve, but they do not change the general learning outcome: Both lexical and grammar learning are successful with this symmetrical singular vs. plural distribution.

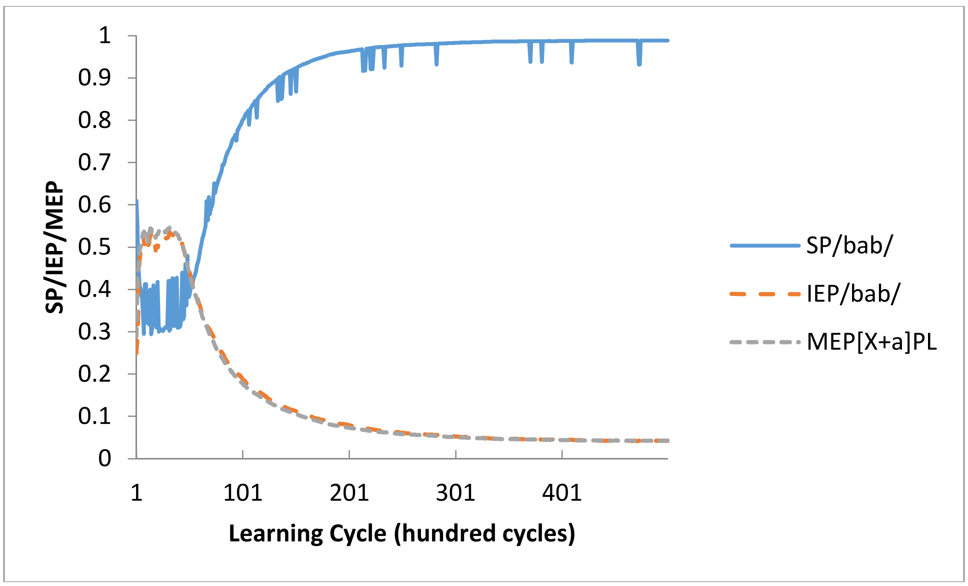


Figure 2.4. Development of $SP_{/bab/}$, $IEP_{/bab/}$, and MEP[X+a]$_{PL}$ with the frequency ratio 1:1

The development significantly differs in Figure 2.5 and Figure 2.6 with a stronger bias toward /bap/. In Figure 2.5, although IEP and MEP plummet as in Figure 2.4, the peak of $SP_{/bab/}$ only barely reaches 0.6 at the 25,000th cycle. In Figure 2.6, the free-falling IEP and MEP after the 10,000th cycle do not produce any significant effect as $SP_{/bab/}$ stays very close to zero. The difference lies in the various biases toward the allomorph /bap/; the stronger is the bias, the lower is $SP_{/bab/}$. If the algorithm continues iterating, it is very likely that the target grammar and output cannot be converged, and if this scenario occurs as a real learning outcome, it means that learners acquire a different lexical and grammar generalization that could eventually lead to diachronic changes.

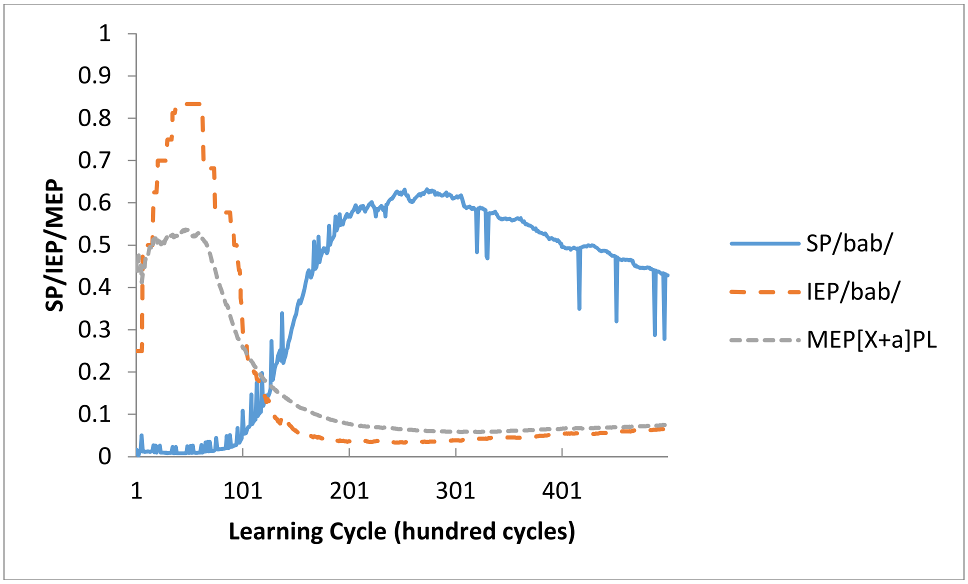


Figure 2.5. Development of $SP_{/bab/}$, $IEP_{/bab/}$, and $MEP[X+a]_{PL}$ with the frequency ratio 9:1

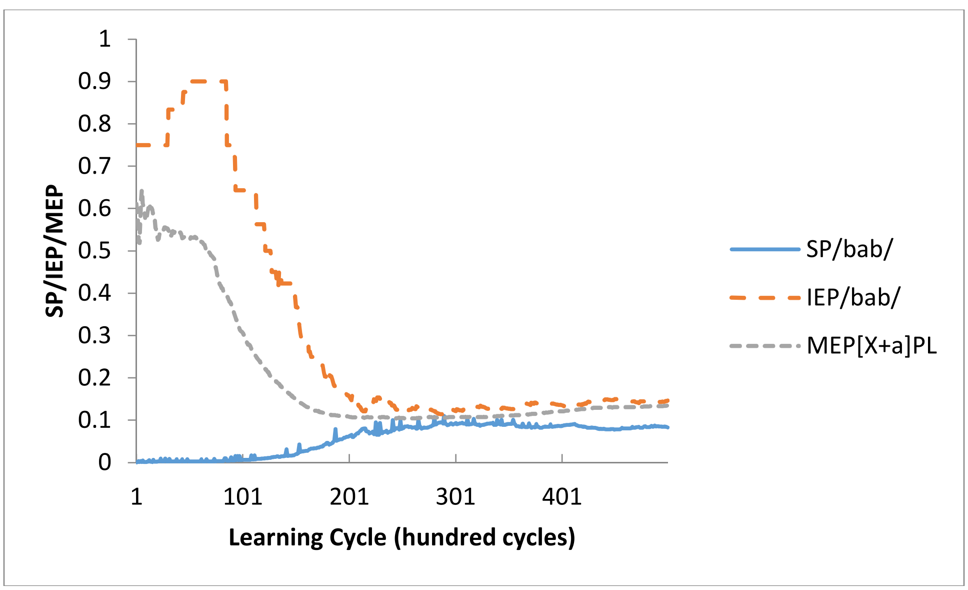


Figure 2.6. Development of $SP_{/bab/}$, $IEP_{/bab/}$, and $MEP[X+a]_{PL}$ with the frequency ratio 19:1

We can explain why $SP_{/bab/}$ with a bias ratio 9:1 cannot be dominant (at least 0.9) by estimating the required $IEP_{/bab/}$ and $MEP[X+a]_{PL}$ in this context using formula (20) and (21). We firstly assume a $freq_{/bap/}$ of nine and a $freq_{/bab/}$ of one, which follow the bias ratio 9:1. $IEP_{/bab/}$ at the 25,000th cycle is selected for the estimation of the $MEP[X+a]_{PL}$ for a dominant $SP_{/bab/}$. Similarly, $MEP[X+a]_{PL}$ at the 25,000th cycle is used to predict the $IEP_{/bab/}$ for a dominant $SP_{/bab/}$. The 25,000th cycle is crucial here as $SP_{/bab/}$ reaches the peak at this point but is not dominant (see Figure 2.5); it is of interest why the lexical variables at this point can produce the highest $SP_{/bab/}$ but still fail to generate a dominant one.

The results in (30) show that either $IEP_{/bab/}$ lowered to 0.007 from 0.03 or $MEP[X+a]_{PL}$ lowered to 0.013 from 0.06 can produce an $SP_{/bab/}$ of 0.9. For the strongest bias 19:1, we anticipate an even lower $IEP_{/bab/}$ and $MEP[X+a]_{PL}$ required for a dominant $SP_{/bab/}$ at the same point. Recall that the input /bab/ may surface as output errors like *[ba] and *[ba.ba] in the very beginning of the M stage simulation, for which $IEP_{/bab/}$ and $MEP[X+a]_{PL}$ cannot be as low as required.

(30) Required $IEP_{/bab/}$ and $MEP_{[X+a]PL}$ for an $SP_{/bab/}$ of 0.9 at the 25,000th cycle with a 9:1 bias ratio

$$ATF_{/bap/} = freq_{/bap/} \times (IEP_{/bap/} \times MEP[X+\emptyset]_{SG})^{-1} = 9 \times (0.13 \times 0.25)^{-1} = 276.9$$

$$IEP(/bab/, SP(A) = 0.9) = \frac{(1-0.9) \times freq_{/bab/}}{0.9 \times ATF_{/bap/}} \times \frac{1}{MEP[X+a]_{PL}} = \frac{(1-0.9) \times 1}{0.9 \times 276.9} \times \frac{1}{0.06} \approx 0.007$$

$$MEP([X+a]_{PL}, SP(A) = 0.9) = \frac{(1-0.9) \times freq_{/bab/}}{0.9 \times ATF_{/bap/}} \times \frac{1}{IEP_{/bab/}} = \frac{(1-0.9) \times 1}{0.9 \times 276.9} \times \frac{1}{0.03} \approx 0.013$$

Whether $SP_{/bab/}$ is dominant or not also affects the end state of the constraint grammar. If $SP_{/bab/}$ is not dominant, the input of Stem-A (pl.) has a higher chance to be /bap+a/, which requires intervocalic voicing in action to generate the target [ba.b-a]; IDENT(voi)-IO can never reliably outrank *V[-voi]V, and grammar learning fails with a bias ratio of 9:1 and 19:1.

5.5 Local summary

The simulation of the toy data has demonstrated how the target grammar can or cannot be achieved with a restrictive constraint re-ranking process and how the target allomorph may or may not be assigned a high SP the combined effect of IEP and MEP. The discussion of intermediate stages were omitted on purpose because the toy data was made up for either a perfectly ideal or an extremely biased learning input distribution, and it may be inappropriate to speculate what really occurs during morphophonological acquisition with the output patterns produced before the end state above. Thus, the discussion of intermediate stages is deferred to Ch. 3, in which the training corpus includes the token frequency of learning inputs in child-directed speech in Dutch, and the outputs at different points of a simulation can be compared to experimental results of different learner groups.

6. Advantages over previous morphophonological learning models

Before moving forward to the case studies of learning a real morphophonology, I will briefly compare PSI-OT-GLA with previous approaches attempting to solve the circularity problem (i.e. a mutual dependence between lexical and grammar learning; see §2) and explain why PSI-OT-GLA is potentially more plausible as a model of morphophonological acquisition.

6.1 Inconsistency detection

The inconsistency detection algorithm in Tesar et al. (2003), Tesar & Prince (2003), Tesar (2004), Merchant & Tesar (2008) tests different UR assumptions by examining their compatibility with a Constraint-Demotion process (e.g. Tesar 1995) which ranks every constraint that prefers target outputs at the top of the ranking hierarchy and generates a single constraint ranking for the target outputs to be optimal. If any UR assumption is incorrect, two or more conflicting constraint rankings may be required for different target outputs. We have seen that in the hypothetical language, deriving [ba.d-a] from /bat+a/ as 'Stem-A (pl.)' requires the ranking *V[-voi]V » IDENT(voi)-IO, which is opposite to the one acquired in the P stage. This ranking inconsistency forces the algorithm to change the UR assumption of 'Stem-A (pl.)' from /bat+a/ to /bad+a/ for the next constraint demotion process, which will be compatible with the ranking *VOICEDOBSCODA » IDENT(voi)-IO » *V[-voi]V.

This learning strategy has a few drawbacks. First, the inconsistency detection algorithm may have a large UR space to search for correct URs; if it is lucky, it might find correct URs within a few iterations, or it might have to rule out a great number of UR candidates before converging on correct ones. If abstract representations are also possible, the search space may grow exponentially. Even if the learning efficiency can be improved in a revised version in Merchant & Tesar (2005), in which eliminating one UR assumption implies eliminating others, the above fundamental issues still remain. Second, without introducing lexical factors, inconsistency detection algorithm is not capable of capturing the effect of token frequency bias, morphological privilege, memory decay, etc. in morphophonological acquisition. In particular, since UR assumptions that result in any ranking inconsistency are simply ruled out, the algorithm disallow relexicalization (learning different surface paradigms (allomorphs) as the basic allomorph of a morpheme; see §2 of Ch. 1, Ch. 4, and Ch. 5).

6.2 Maximum Likelihood Learning

Jarosz (2006b, 2007, 2011) proposes Maximum Likelihood Learning of Lexicons and Grammars as an algorithm that exhaustively computes the probabilities distributed across a set of possible UR assumptions and different constraint grammars generated by *n* constraints, and the learning outcome would be the UR and grammar probabilities that can generate output patterns which best approximate learning inputs. With the four surface forms in the hypothetical example, the best UR probability distribution is presumably be 2% for the UR /bap/ and 98% for /bab/, and the best grammar probability distribution could be 97% for the ranking *VOICEDOBSCODA » IDENT(voi)-IO » *V[-voi]V (for the majority of correct derivations with final devoicing; e.g. /bab/→[bap]), 2.8% for the ranking *VOICEDOBSCODA » *V[-voi]V » IDENT(voi)-IO (for some correct derivations that require intervocalic voicing; e.g. /bap+a/→[ba.b-a]), and 0.2% for the rest of possible rankings. The two probability distributions shed light on what combination of a UR assumption and a constraint ranking best fits the learning data (i.e. UR = /bab/ and ranking = *VOICEDCODA » IDENT(voi)-IO » *V[-voi]V). The circularity problem is solved in this algorithm since the acquisition order of the lexicon and the grammar does not matter; the algorithm simply tests every possible combination and concludes with the most robust model.

The goal of solving the circularity problem is achieved at the considerable cost of the computational efficiency, however, since the number of the probability combinations soars as the number of URs and constraints increase (Albright & Hayes 2011:671). It is possible to slightly limit the computational space of URs by including only surface-true allomorphs (see §1.1 as well), but the number of possible constraint grammars generated by *n* innate constraints (i.e. *n!*) could be incredibly large. If we include learnable unnatural constraints as well, the learning process in

Maximum Likelihood Grammar may never complete as the number of testable grammars keeps growing. In short, while the algorithm can capture some output patterns in morphophonological acquisition, the learning strategy might be too effortful to be implemented by real learners.

In addition to the computational burden, the algorithm does not directly incorporate lexical factors into the model, thus missing some crucial developments in morphophonological acquisition. It can partially account for the correlation between the token frequency of a form with a morphophonemic alternation and the time necessary for acquiring the form as modeled in Jarosz (2011):[29] Learners converge on the correct UR of morphologically complex forms with a higher token frequency since these forms have a higher chance of being produced for their UR assumptions to be tested (see §1.1 and §1.2). Nevertheless, the algorithm does not introduce token frequency as a principle factor of determining a basic allomorph and is unable to capture output variation influenced by a token frequency bias toward some allomorphs, and in this regard relexicalization may not occur either. A less precise modeling in morphophonological acquisition is also expected with the absence of MEP and memory decay.

6.3 Lexical constraint

Another attempt to solve the circularity problem is to introduce learnable lexical constraints into the constraint grammar, which prohibit an allomorph to be the UR of a morpheme; selecting /bap/ as the UR of 'Stem-A' violates */bap/→Stem-A and selecting /bab/ violates */bab/→Stem-A, and the ultimate choice is determined upon the ranking between the two. This approach was proposed with the GLA in Apoussidou (2006) and a batch-learning Maximum Entropy model in Eisenstat (2009) and Pater et al. (2012). For the UR of 'Stem-A' to be /bab/, the final grammar should have */bab/→Stem-A being outranked by */bap/→Stem-A.

In this approach, candidates of an evaluation process are no longer possible outputs but a set of possible input-output mappings. The winner becomes the mappings that violates least high-ranked constraints. The circularity problem is

[29] In her modeling of the final voicing alternation in Dutch, Jarosz (2011) in fact did not attempt to account for the developmental difference between individual forms. Instead, she was comparing the acquisition of a morphological 'type' to the acquisition of other types. That is, Dutch plural forms with the voicing alternation has the lowest token frequency if compared to Dutch singular forms and plural forms without the voicing alternation, and the low token frequency explains why Dutch plural forms are acquired slower 'in general'.

avoided as the correct URs and target grammar are acquired simultaneously via the same constraint promotion and demotion process. In its incipient stage, the algorithm will consistently rank IDENT(voi)-IO higher than *V[-voi]V because the intervocalic voicing contrast is preserved in [ba.t-a] 'Stem-B (pl.)'. If */bap/→Stem-A is also ranked lower than */bab/→Stem-A at this stage, /bap+a/ must be the input of 'Stem-A (pl.)', which can only be mapped faithfully to an output error as in Tableau 2.4. Such an input-output mapping triggers the demotion of */bab/→Stem-A since only the input /bab+a/ can surface as the correct output with the constraint ranking IDENT(voi)-IO » *V[-voi]V. Eventually, the grammar should be stabilized with the ranking {*/bap/→Stem-A, *VOICEDOBSCODA} » IDENT(voi)-IO » {*V[-voi]V, */bab/→Stem-A} which can generate every surface target form, and by this ranking the UR of Stem-A is determined (i.e. /bab/).

	*/bab/→Stem-A	IDENT(voi)-IO	*/bap/→Stem-A	*V[-voi]V
☞☹/bap+a/→[ba.p-a]			←*	←*
✓/bab+a/→[ba.b-a]	*!→	*→		

Tableau 2.4. Constraint re-ranking triggered by an input-output mapping error; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

The main advantages are twofold. First, the algorithm does not have to try different UR assumptions one at a time as in inconsistency detection; the algorithm evaluates UR assumptions and acquires phonotactic knowledge at the same time in a gradual constraint promotion and demotion process. Second, in a stochastic constraint grammar, */bap/→Stem-A may still be probabilistically outrank */bab/→Stem-A to produce output variations as a result of the probabilistic input selection. In addition, lexical factors in PSI-OT-GLA can be easily introduced into the lexical constraint framework as well. First of all, the token frequency of each allomorph can be introduced as a ranking variable that changes the constraint value of a corresponding lexical constraint (à la Coetzee & Kawahara 2013). Lexical constraints corresponding to high-frequency allomorphs are thus ranked lower than indicated by their original constraint value, and those indexed to low-frequency allomorphs are ranked higher than expected;[30] input-output mappings with high-frequency allomorphs IEP is also reflected in the constraint value of

[30] See Ch. 6 for an approach that associates a token frequency to a number of violation marks of a lexically indexed faithfulness constraint.

lexical constraints per se; the number of output errors generated with an allomorph is equal to the frequency of promoting the corresponding lexical constraint, whereas the number of correct outputs is equal to the frequency of demoting the constraint. Although MEP and memory decay are still absent, the two principle factors are expected to make predictions similar to those in PSI-OT-GLA, such as frequency bias, word-specific input preference, and relexicalization.

However, input selection in this framework is only possible when lexical constraints are not dominated by markedness and faithfulness constraints. Consider the following case: If *V[-voi]V is accidentally higher-ranked than both */bab/→Stem-A and */bap/→Stem-A, and IDENT(voi)-IO also outranks *V[-voi]V to preserve the intervocalic voicing contrast, the optimal mapping of 'Stem-A (pl.)' is always /bab+a/→[ba.b-a] regardless of the ranking between */bab/→Stem-A and */bap/→Stem-A as in Tableau 2.5.

	IDENT(voi)-IO	*V[-voi]V	*/bab/→Stem-A	*/bap/→Stem-A
☞/bab+a/→[ba.b-a]			*	
/bap+a/→[ba.b-a]	*!			*
/bap+a/→[ba.p-a]		*!		*

Tableau 2.5. No effect of lexical constraints with dominant markedness and faithfulness constraints

In addition, since lexical constraints indexed to high-frequency allomorphs are assumed to be further demoted, they have a lower chance to probabilistically outrank markedness and faithfulness constraints to force input selection than those indexed to low-frequency allomorphs, which is also an unwanted result.

To constantly take token frequency into account, it is necessary for a model to consider all violation marks of a candidate during the evaluation process as permitted in Harmonic Grammar. This alternative model will be discussed as an expansion of PSI-OT-GLA in Ch. 6, and this section simply demonstrates why the primitive lexical constraint framework is inadequate if compared to PSI-OT-GLA.

6.4 Summary

To summarize, I consider PSI-OT-GLA to be a morphophonological acquisition model which parallels previous models in providing a formal account of morphophonological acquisition. PSI-OT-GLA nevertheless is expected to outperform these models in terms of lexical learning by constantly incorporating a set of lexical factors that can influence the selection of basic allomorphs at the input level. In the next few chapters, we will see how this model accounts for output patterns in morphophonological acquisition and diachronic morphophonemic changes as the result of gradual shifts in selecting stored allomorphs as inputs.

Chapter 3

Probabilistic Selection of Input and the acquisition of Dutch final devoicing

In this chapter, PSI-OT-GLA is implemented to capture different stages in the acquisition of a stem-final voicing alternation in Dutch. The data in (1) includes morphologically related forms with a voicing alternating and non-alternating stem respectively. As in hypothetical language constructed in Ch. 1, the target constraint grammar *VOICEDOBSCODA » IDENT(voi)-IO » *V[-voi]V forbids any voiced obstruent coda but maintains the intervocalic voicing contrast, and the target allomorphs with a dominant SP are plural allomorphs. Not only will we see below how this end state can be achieved in PSI-OT-GLA, but also how the transitional stages reported in Kerkhoff's (2007) Experiment I can be approximated with the proposed learning model.

(1) Final devoicing in Dutch

a.	[bɛt]	'bed'	[bɛ.d-ən]	'beds' (alternating stem)
b.	[pɛt]	'cap'	[pɛ.t-ən]	'caps' (non-alternating stem)

1. Children's production of Dutch voicing alternation

This chapter begins with the discussion of Kerkhoff's (2007) experimental study of the morphophonological development of Dutch-learning children using the Wug test paradigm (Berko 1958). Kerkhoff's study specifically focuses on the acquisition of the stem-final voicing alternation presented in (1) and examines the output patterns collected in her experiment. This section summarizes the results of Experiment I in Kerkhoff's study, which asked children participants to produce real Dutch plural forms. The summary aims to further specify the output patterns that a simulation based on PSI-OT-GLA needs to replicate.

In her Experiment I, Kerkhoff presented a picture of a single object corresponding to a real Dutch singular form to Dutch-learning children in each trial. The children were then encouraged to practice the singular form until they produced it correctly. A picture showing the same but multiple objects was then presented to the participant to elicit the corresponding plural form. Sixteen target plural forms included in the elicitation task are listed in Table 3.1: Eight plural forms like [bɛ.d-ən] 'beds' whose stem-final obstruent is devoiced in their singular form are called alternating plural forms (AP), and another eight plural forms without the stem-final voicing alternation across different contexts, such as [pɛ.t-ən] 'caps', are non-alternating plural forms (NAP). Token frequencies are obtained from the CELEX corpus (Baayen et al. 1995).

NAP	Target	Token freq.	AP	Target	Token freq.
caps	[pɛtən]	2	beds	[bɛdən]	12
foots	[vutən]	129	turtles	[sxilpɑdən]	2
tents	[tɛntən]	7	hats	[hudən]	4
elephants	[olifɑntən]	4	pencils	[pɔtlodən]	2
chickens	[kɪpən]	14	hands	[hɑndən]	377
monkeys	[ɑpən]	9	dogs	[hɔndən]	53
sheeps	[sxapən]	15	webs	[wɛbən]	0
lamps	[lɑmpən]	10	crabs	[krɑbən]	0

Table 3.1. Non-alternating (NAP) and alternating (AP) Dutch plural forms with their target output and token frequency

Three age groups of Dutch learning children were involved in the experiment (i.e. 24 children of 2;9-3;11, nineteen children of 4;0-6;2, and fifteen children of 6;9-7;8), which could help demonstrate the chronological development of Dutch final devoicing. There are various output error types of plural forms, but we simply focus on the major intervocalic (de)voicing error; that is, APs produced with a voiceless stem-final obstruent (e.g. *[bɛ.t-ən]) and NAPs produced with a voiced stem-final obstruent (e.g. *[pɛ.d-ən]). This is not saying the cause of the minor error types are not worth noting, but that they might not be the result predicted by the input selection process. These output error types include bare stems (i.e. produced without a plural suffix; 24 tokens of APs and 28 tokens of NAPs), stem segment changes (five tokens of APs and none of NAPs), S-plural (i.e. produced with /-s/ suffix; five tokens of APs and six tokens of NAPs), and missing (i.e. no response; 52 tokens of APs and 43 tokens of NAPs).[31] After excluding the number of minor error types (86 tokens = 18.6% of the AP results and 77 tokens = 15.9% of the NAP results),

[31] See Table 21 and 22 in Kerkhoff (2007:149-150).

the output error rates are calculated using formulae (2) in Table 3.2.

(2) Calculation of voicing error rates

Alternating plural form error rate = $\frac{intervocalic\ devoicing\ errors}{intervocalic\ devoicing\ errors+correct\ outputs}$

Non-alternating plural form error rate = $\frac{intervocalic\ voicing\ errors}{intervocalic\ voicing\ errors+correct\ outputs}$

AP	2;9-3;11	4;0-6;2	6;9-7;8
Devoicing Error	61% (81)	49.6% (63)	41.5% (49)
Correct Output	39% (52)	50.4% (64)	58.5%(69)
NAP			
Voicing Error	3.8% (6)	5.4% (7)	1.7% (2)
Correct Output	96.2% (152)	94.6% (122)	98.3% (118)

Table 3.2. Output type rates and token numbers of alternating plural forms (AP) and non-alternating plural forms (NAP)

One interesting observation that can be made from these results is the correlation between the improvement in the production of APs and the increasing voicing errors of NAPs. Specifically, from the youngest group to the 4;0-6;2 group, the chance of NAP voicing errors increases as the percentage of AP devoicing errors drops. After this stage, both error rates drop to the lowest point in the oldest group. Of course, the growing rate of NAP voicing errors between the two younger groups might be debatable due to the scarcity of these errors. However, the results of individual subjects from different age groups show that most NAP errors were produced by those who produced more correct AP outputs (Figure 3 in Kerkhoff 2007:154). In particular, one subject produced all target APs correctly but also produced nearly 40% of target NAPs as intervocalic voicing errors. We thus follow Kerkhoff's conclusion that the growing NAP error rate is closely correlated with the decreasing AP error rate in the rest of this chapter.

Kerkhoff (2007:152, Figure 2) also discusses the output pattern of each AP demonstrated by all three age groups, which is re-calculated with formula (2) in Table 3.3.[32] One apparent trend in this data set is that token frequencies are in

[32] I am grateful to Annemarie Kerkhoff (p.c., Sep. 2013) who generously shared the detailed results with me, which were originally excluded from Figure 2 in her dissertation work.

general related to devoicing error rates. High-frequency APs like ‘hands’ and ‘dogs’ can have an error rate lower than 30%. Low-frequency APs like ‘turtles’ and ‘hats’ have a much higher error rate, and the zero-frequency of APs ‘webs’ and ‘crabs’, not surprisingly, leads to top error rates. The exception that can be singled out is the AP ‘pencils’, which has an error rate similar to that of high-frequency APs while having an extremely low token frequency.

AP	Target	Token freq.	Devoicing Error
beds	[bɛdən]	12	46% (23)
turtles	[sxilpɑdən]	2	69% (33)
hats	[hudən]	4	73% (32)
pencils	[pɔtlodən]	2	26% (12)
hands	[hɑndən]	377	24% (12)
dogs	[hɔndən]	53	18% (9)
webs	[wɛbən]	0	78% (32)
crabs	[krɑbən]	0	82% (40)

Table 3.3. Devoicing error rates and total error tokens of individual alternating plural forms in the elicitation task

Considering the above discussion, a successful PSI simulation must include the acquisition of the final devoicing grammar and produce the output patterns documented in Kerkhoff (2007). Accordingly, PSI will be evaluated with the checklist (3), in which the primary modeling goals are specified.

(3) Goals of modeling Dutch final devoicing with PSI

a. A much higher NAP voicing error rate than the AP devoicing error rate in all three stages.
b. An intermediate stage in which the NAP error rate grows while the AP error rate drops.
c. A higher error rate for low-frequency alternating plural forms.
d. A constraint ranking of final devoicing.

2. Training corpus and constraint set

This section discusses individual components in the simulation of the production performance of the Dutch stem-final voicing alternation, which include a training corpus composed of the same sixteen plural forms as well as their singular counterparts (i.e. 32 forms in total), the structure of learning inputs, and a set of crucial constraints.

2.1 Training input distribution

The 32 items in the training corpus are listed in Table 3.4 with their individual raw token frequency in CELEX (Baayen et. al 1995), which is converted into a distributional probability. These raw token frequencies in adult speech were reported to be reminiscent of child-directed speech (CDS) in Kerkhoff (2007, §4.3 of Ch. 4). In the computer simulations below, one of the 32 forms will be randomly fed to the algorithm in every learning cycle, and the distribution probabilities in Table 3.4 determine the chance of 'perceiving' an input by the algorithm.

By storing each input and tracking its token frequency, the distributional probabilities in the lexicon are expected to approach those in the training corpus after thousands of learning cycles, and the algorithm should thus 'produce' the same word (i.e. check whether a word can surface as its correct output) with a similar probability, which is in accord with the observation that the frequency distribution in CDS and that in child speech are highly correlated (Kerkhoff 2007:§4.4.4.1). In sum, in the following simulations, one input is randomly 'perceived' and 'produced' by the algorithm. The probability of perceiving and producing each item is determined by the distributional probabilities in the training corpus and the lexicon in each learning cycle.

NAP Stem	singular	freq sg.	plural	freq pl.	sg-pl ratio
cap	[pɛt]	16 (0.83%)	[pɛtən]	2 (0.1%)	8:1
foot'	[vut]	96 (4.98%)	[vutən]	129 (6.69%)	0.74:1
tent	[tɛnt]	20 (1.04%)	[tɛntən]	7 (0.36%)	2.86:1
elephant	[olifɑnt]	6 (0.31%)	[olifɑntən]	4 (0.21%)	1.5:1
chicken	[kɪp]	19 (0.99%)	[kɪpən]	14 (0.73%)	1.36:1
monkey	[ɑp]	12 (0.62%)	[ɑpən]	9 (0.47%)	1.33:1
sheep	[sxap]	11 (0.57%)	[sxapən]	15 (0.78%)	0.73:1
lamp	[lɑmp]	21 (1.09%)	[lɑmpən]	10 (0.52%)	2.1:1
AP Stem					
bed	[bɛt]	284 (14.74%)	[bɛdən]	12 (0.62%)	23.66:1
turtle	[sxilpɑt]	4 (0.21%)	[sxilpɑdən]	2 (0.1%)	2:1
hat	[hut]	31 (1.61%)	[hudən]	4 (0.21%)	7.75:1
pencil	[pɔtlot]	10 (0.52%)	[pɔtlodən]	2 (0.1%)	5:1
hand	[hɑnt]	645 (33.47%)	[hɑndən]	377 (19.56%)	1.71:1
dog	[hɔnt]	107 (5.55%)	[hɔndən]	53 (2.75%)	2.02:1
web	[wɛp]	3 (0.16%)	[wɛbən]	0 (0%)	--[33]
crab	[krɑp]	2 (0.1%)	[krɑbən]	0 (0%)	--

Table 3.4. Token frequencies, distributional probabilities, and singular-plural ratios of the 32 Dutch forms from CELEX; AP = alternating plural form, NAP = non-alternating plural form

There are a few other distributional properties that might affect following simulations. First, the current input distribution in Table 3.4 is clearly skewed on two dimensions: Alternating forms are more frequent than non-alternating forms, and singular forms are more frequent than plural forms.[34] The second frequency bias (represented by 'sg-pl ratio' in Table 3.4) crucially predicts that the singular allomorphs will benefit from their high token frequency in the competition against the plural allomorphs in the input selection process. This is of course not saying that low frequency plural allomorphs always lose the battle; as demonstrated in §5 of Ch. 2, in some conditions with a low IEP and MEP, the low frequency plural allomorphs can still gradually overcome the disadvantage, and we will see such development later in this chapter.

[33] The bias ratio cannot be calculated with a zero token frequency of 'webs' and 'crabs' in the CELEX database.

[34] Only two non-alternating pairs [vut]~[vutən] and [sxap]~[sxapən] have a minute frequency bias toward the singular form.

Second, among these 32 forms in the training corpus, the APs 'webs' and 'crabs' have a token frequency of zero, which rules out the possibility for them to be either perceived or produced anytime during morphophonological acquisition modeled with PSI-OT-GLA. Therefore, during a task of producing these APs in an elicitation test, the algorithm predicts that the plural allomorph of the alternating stems can never be accessed.[35]

Finally, a few items in the training corpus have an extremely low token frequency and distributional probability like 'pets' (0.1%), 'elephants' (0.21%), 'turtle' (0.21%), 'turtles' (0.1%), 'pencils' (0.1%), 'web' (0.16%), and 'crab' (0.1%). The following simulations also assume that forms with an extremely low token frequency are lexically unstable and highly vulnerable to the memory decay effect.[36] The decay rate was thus set as 0.002 per learning cycle (or 2 token per 1,000 learning cycles), which may have at times erased these forms from the lexicon. Similar to those intrinsically absent from the training corpus, the forms completely decayed at any given point have a zero token frequency and the stem allomorphs in these forms (e.g. /sxilpɑd/ from [sxilpɑdən] and /pɔtlod/ from [pɔtlodən]) cannot be accessed in the input selection process.

2.2 Structure of inputs and target outputs

Since our only focus here is the acquisition of the final devoicing alternation in Dutch, the structure of the input representation is considerably simplified. The inputs and the targets similar to those in the hypothetical data are listed in (4) will be used to represent each morphophonemic alternation type. For example, the input of alternating singular forms is always either /bɛt/ or /bɛd/, and the target is always [bɛt]. For non-alternating stems, the input is always /ɑp/ and the target output is always an onsetless syllable as in [ɑp] and [ɑ.p-ən].

[35] This is not saying that the alternation can never be produced without accessing the allomorph. See the discussion in the rest of the chapter.

[36] This of course does not suggest that other low frequency forms such as [olifɑnt], [olifɑntən], etc. are always stable in the lexicon. The decay rate was assumed more conservatively to avoid eliminating too many forms from the lexicon and ultimately undermining the morphophonological learning process with an 'over-shrinking' lexicon.

(4) Inputs and outputs of different morphological types in the Dutch simulation

a. Alternating singular form – input: /bɛt/ or /bɛd/, target: [bɛt]

b. Alternating plural form – input: /bɛt+ən/ or /bɛd+ən/, target: [bɛ.d-ən]

c. Non-alternating singular form – input: /ɑp/, target: [ɑp]

d. Non-alternating plural form – input: /ɑp+ən/, target: [ɑ.p-ən]

This input set differs from the hypothetical data used in Ch. 2 in the sense that all the target outputs in the latter have an onset and leads to a top-ranked ONSET constraint in the acquired grammar. The reasons for the change are twofold. The first one is the empirical fact in Dutch phonotactics learning. Levelt et al. (1999) nevertheless report that onsetless syllables can be acquired as early as 1;6 – an age much younger than the threshold for mastering the voicing alternation in the Dutch morphophonology (see below). Therefore, the simulation below must capture this development in phonotactic learning, and ONSET should be ranked at the bottom by the end of the following simulations. Note that onset clusters in alternating and non-alternating forms like [sx] in [sxilpɑt] are excluded from the inputs to reduce the number of possible outputs and accelerate computer simulation. The onset clusters, like onsetless syllables, are acquired in the phonotactic stage with a lower-ranked *COMPLEX as reported in Levelt et al. (1999) and are unlikely to affect morphophonological learning. The second reason is algorithmic. In §5.3 of Ch. 2, I have demonstrated that if ONSET is top-ranked, there is no need to promote MAX-IO over *VOICEDOBS since the deletion of a voiced obstruent (e.g. /bɛt/→*[ɛt]) during phonotactics learning is forbidden by ONSET. Since the ranking of faithfulness constraints affects the acquisition of final devoicing, which requires IDENT(voi)-IO to dominate MAX-IO and DEP-IO, we certainly hope the promotion and demotion of a constraint in the following simulation are based on the input types that can be observed by real learners; the development in the simulation can thus best approximate real learners' performance.

2.3 Constraint set

The same eight constraints in the toy data simulation in §2.5 will be included in the following simulations as well, which are repeated as (5). The initial and target ranking are summarized in (6).

(5) Constraints for grammar learning in Dutch morphophonological acquisition

a. ONSET: Syllables without onset are prohibited.
b. *VOICEDOBS (*VOIOBS): Voiced obstruents are prohibited.
c. *VOICEDOBSCODA (*VOIOBSCODA): Voiced obstruent codas are prohibited.
d. *CODA: Syllable codas are prohibited.
e. *V[-voi]V: Intervocalic voiceless consonants are prohibited
f. MAX-IO: Every input segment must have an output correspondent.
g. DEP-IO: Every output segment must have an input correspondent.
h. IDENT(voi)-IO: Every input specification of [voice] must be preserved in the output.

(6) Initial and target constraint ranking in Dutch morphophonological acquisition

Initial Ranking: {ONSET, *VOICEDOBS, *VOICEDOBSCODA, *CODA, *V[-voi]V} » {MAX-IO, DEP-IO, IDENT(voi)-IO}

Target Ranking: {*VOICEDOBSCODA, MAX-IO, DEP-IO} » IDENT(voi)-IO » {ONSET, *VOICEDOBS, *CODA, *V[-voi]V}

Note that the unnatural constraint *V[+voi]V (i.e. intervocalic devoicing) is excluded from the discussion here. In the training corpus, there are fourteen plural forms in total (excluding the zero frequency forms 'webs' and 'crabs'), and based on the Tolerance Principle (see §2.2.1 of Ch. 2), the unnatural constraint can be acquired only when the number of exceptions is lowered than 16 / log(16) = 5.7. Among the sixteen plural forms, there are eight NAPs as the exceptions to *V[+voi]V, which is higher than the tolerance threshold.[37] The unnatural constraint thus is not expected to be created with the training corpus. It does not, however, rule out the possibility for the constraint to be created with a larger corpus in which NAPs greatly outnumber APs.

The eight constraints allow us to posit possible outputs allowed in the GEN(ERATOR) function in OT for each of the six inputs in (4), and following Riggle's (2004) Finite-State OT model, PSI-OT-GLA assumes that non-harmonically bounded outputs will be generated from each input. Readers are referred to

[37] Based on the CELEX corpus counting (Kerkhoff 2007:103), the type frequency of alternating plural forms is 313 out of a total of 1,234 plural types. That is, there are 313 exceptions to intervocalic devoicing, which is higher than the threshold (1,234 / log(1,234) = 173.4). Therefore, even if we assume that learners have the chance to perceive all plural forms, the unnatural markedness constraint *V[+voi]V is unlikely to be created by the learners.

Appendix B for a full list of Elementary Ranking Conditions (ERC; Prince 2002). Finally, although the input representation of the same morphological type is always identical, each form in Table 5 still has its own token frequency and each allomorph in an alternating pair still has an individual SP and IEP, i.e. the probability of selecting /bɛt/ and /bɛd/ as the input is different for each alternating stem.

3. Elicitation task in computer simulation

A valid comparison between Kerkhoff's (2007) experimental results and the following simulated results requires mimicking the settings in Kerkhoff's experiment in PSI-OT-GLA. There are three settings in Kerkhoff's elicitation task (Experiment I) crucial to setting up similar simulations below, which resembles those dealing with essentially the same data in Jarosz (2011). First, the three different age groups participating Experiment I in Kerkhoff (2007) (2;9-3;11, 4;0-6;2, and 6;9-7;8) cannot be directly converted into the number of learning cycles in the M stage simulation. In a few pre-tests, it was shown that distinct learning phases are observable within 100,000 learning cycles in the M stage simulation. The simulated elicitation task was thus conducted for every fifty learning cycles within this number of iteration. The results generated in these cycles are then compared with Kerkhoff's experimental results to evaluate the correspondence between learning cycles and age groups. Second, the participants in Kerkhoff's experiment had to name a picture of a singular object in each trial and were encouraged to repeat the singular form until the form is produced correctly. The correct naming of a singular object means that the token frequency of the corresponding singular form must not be zero immediately before the elicitation task. In the current simulation, the token frequency of all singular forms thus undergoes add-one smoothing before every simulated elicitation (i.e. add one to the raw token frequency of each singular form for every 50 learning cycles). Finally, to simulate the elicitation process (i.e. the subjects were asked to name the plural form of the object), each of the APs and non-APs were 'elicited' from the machine learner for 100 times after every 50 learning cycles to obtain an averaged performance with the constraint values and lexical factors developed up to this point, and the results will be discussed in the following sections.

4. Simulated results

The numbers of learning cycles (i.e. 50,000 in the P stage simulation and 100,000 in the M stage simulation) were determined by pre-tests in which the expected learning outcome was found. The simulation was executed 100 times, so constraint values, SPs, IEPs, and MEPs were numbers averaged over 100 learning outcomes. The goal is to demonstrate the typical development of the singular and plural forms in PSI-OT-GLA, instead of one single pattern which matches children's performance in Kerkhoff's experiments merely by accident. Likewise, the simulation was compiled in a Java® environment in Eclipse Standard version 4.3.2 (The Eclipse Foundation 2014).[38]

4.1 Phonotactic learning

I again start with the acquisition of phonotactic knowledge, which is the first half of the two-stage learning process and only adjusts constraint values. Since the input selection process is inactive in this stage, the input is always identical to the target and thus only four inputs representing each morphological type (i.e. /bɛt/, /bɛdən/, /ɑp/, and /ɑpən/) are submitted to the algorithm with their ERCs. The constraint ranking translated from the adjusted constraint values at the end of the P stage simulation is illustrated in Table 3.5. As in §5 of Ch. 2, when two constraints are separate with a constraint value difference of 5, the one with a lower value has an extremely low chance to outrank the one with a higher value with the grammar learning settings, and the ranking between the two constraints is represented with strict domination.

	Ranking: *VOIOBSCODA» {IDENT(voi), DEP} » {MAX, ONSET} » {*V[-voi]V*, *VOIOBS} » *CODA
Value	100 53 52.7 48.5 43.8 40.6 38.3 25.2

Table 3.5. Constraint values/rankings acquired at the end of the P stage simulation

The discussion of grammar learning begins with the change from the initial Markedness » IO-Faithfulness ranking to the above interim ranking. First of all, *VOICEDOBSCODA remains top-ranked without any voiced obstruent coda in the target outputs, but IDENT(voi)-IO must be promoted over *VOICEDOBS to prevent voiced obstruents from devoicing in a non-coda position in the inputs /bɛt/ and

[38] The source code is available at http://hdl.handle.net/10402/era.39159.

/bɛdən/, such as in Tableau 3.1. IDENT(voi)-IO is also promoted to outrank *V[-voi]V to preserve the intervocalic voicing contrast as in the input /ɑpən/ (see Tableau 3.2). Due to a higher promotion frequency, IDENT(voi)-IO ends up having the second highest constraint value despite no crucial ranking between IDENT(voi)-IO and ONSET.[39]

/bɛt/	*VOICEDOBS	IDENT(voi)-IO
✓bɛt	*!→	
☞☹pɛt		←*

Tableau 3.1. Promotion of IDENT(voi)-IO to preserve input voiced obstruents; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

/ɑpən/	*V[-voi]V	IDENT(voi)-IO
✓ɑ.pən	*!→	
☞☹ɑ.bən		←*

Tableau 3.2. Promotion of IDENT(voi)-IO to outrank *V[-voi]V; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

The goal of the promotion of DEP-IO is twofold as shown in Tableau 3.3. First, the faithfulness constraint must outrank ONSET to avoid onset insertion for onsetless syllables in inputs /ɑp/ and /ɑpən/. Second, it also has to dominate *CODA to prohibit vowel insertion which changes coda consonants in all four inputs into onsets.

/ɑp/	ONSET	*CODA	DEP-IO
✓ɑp	*!→	*→	
☞☹tɑp		*	←*
☞☹ɑ.pi	*		←*

Tableau 3.3. Promotion of DEP-IO to prohibit the insertion repair to marked syllable structures; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

[39] In the training data, the alternating singular forms have a higher token frequency, and thus the input like /ɑp/ is produced most frequently by the machine learner in the P stage simulation. The ERCs in Appendix B show that when the input is /bɛt/, IDENT(voi)-IO is winner-preferring for 5 out of 8 possible losers (cf. 4/8 for DEP-IO and 3/8 for MAX-IO), which has a highest chance to be promoted when an error output surfaces.

Finally, MAX-IO must dominate all markedness constraints but *VOICEDOBSCODA to avoid deleting voiced obstruents (Tableau 3.4), intervocalic voiceless obstruents, and coda consonants (Tableau 3.5). Note that DEP-IO is promoted slightly over MAX-IO not because the algorithm prefers deletion over insertion in choosing a repair strategy. DEP-IO is simply promoted more often since in the ERCs created for each input in the P stage simulation (see Appendix B), DEP-IO is a winner-preferring constraint more frequently when a non-target output candidate is selected as the optimal output.

/bɛt/	*VOICEDOBS	MAX-IO
✓bɛt	*!→	
☞☹ɛt		←*

Tableau 3.4. MAX-IO is promoted over *VOICEDOBS to preserve voiced obstruents; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

/ɑpən/	*V[-voi]V	*CODA	MAX-IO
✓ɑ.pən	*!→	*	
☞☹ɑ.pə	*		←*
☞☹ɑ.ən		*	←*

Tableau 3.5. MAX-IO is promoted over *V[-voi]V and *CODA to avoid deletion; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

The P stage constraint ranking differs from the target ranking repeated in (7) with the sub-ranking *VOICEDOBSCODA » {IDENT(voi)-IO, DEP-IO} » MAX-IO, which allows the deletion, rather than devoicing, of an underlying voiced obstruent coda. We have seen an undominant IDENT(voi)-IO in §5 of Ch. 2 as well for the same reason: In the P stage simulation, while the lack of any voiced obstruent coda in the training data allows a top-ranked *VOICEDOBSCODA, the correct repair strategy of final devoicing cannot be acquired as there is no potential voiced obstruent coda in the input like /bɛd/ in Tableau 3.6. This is not saying that the simulation fails, but suggesting an intermediate stage which affects the morphophonological acquisition in the M stage simulation below.

(7) Target ranking
{*VOICEDOBSCODA, MAX-IO, DEP-IO} » IDENT(voi) »
{ONSET, *VOICEDOBS, *CODA, *V[-voi]V}

/bɛd/	*VOIOBSCODA	IDENT(voi)	DEP-IO	MAX-IO
✓bɛt		*!		
bɛd	*!			
bɛ.di			*!	
☞☹bɛ				*

Tableau 3.6. Deleting an underlying voiced obstruent coda with the P stage ranking; '✓' = correct output, '☹' = output error

4.2 Morphophonological learning: Prediction

Prior to the discussion of the results in the M stage simulation, I would like review the factors that could influence the learning outcome below and make several predictions to help explain the output patterns and changes in the lexical variables and constraint values. These predictions will correspond to three individual learning phases in the M stage simulation as I spell out below.

Phase I:

Grammar learning:

*VOICEDOBSCODA» {IDENT(voi)-IO, DEP-IO } » {MAX-IO, ONSET} » {*V[-voi]V*, *VOICEDOBS} » *CODA

(the ranking from the P stage simulation, and final devoicing has not been acquired)

Lexical learning:

Singular allomorphs have a higher SP due to a higher raw token frequency.

Output patterns:

/bɛt/→[bɛt] (correct output)

/bɛd/→*[bɛ.di] or *[bɛ] (repairing a voiced coda with deletion or insertion)

/bɛt+ən/→*[bɛ.t-ən] (singular allomorph inputs with a dominant IDENT(voi))

/bɛd+ən/→[bɛ.d-ən] (correct output)

/ɑp/→[ɑp] (correct output)

/ɑp+ən/→[ɑ.p-ən] (correct output with a dominant IDENT(voi))

At the very beginning of the M Stage simulation, which is Phase I, forms with a non-alternating stem should be produced correctly since the faithful mapping between their single possible input and the correct output is allowed by the constraint grammar inherited from the P stage simulation. The input of alternating stems, however, is determined by input selection, and singular allomorphs of alternating stem should have a higher SP, thanks to the frequency bias toward these singular allomorphs as indicated in §2.1. Therefore, the inputs of APs and NAPs are more likely /bɛt/ and /bɛt+ən/ respectively. With the constraint ranking inherited from the P stage simulation, the singular form like [bɛt] can be successfully derived from /bɛt/ as in Tableau 3.7.

/bɛt/	DEP-IO	MAX-IO	*CODA
☞✓bɛt			*
bɛ		*!	
bɛ.ti	*		

Tableau 3.7. Faithful production of /bɛt/

The input /bɛt+ən/, however, cannot surface as [bɛ.d-ən] in Tableau 3.8 because of the ranking IDENT(voi)-IO » *V[-voi]V. Errors like *[bɛ.t-ən] will trigger the algorithm to move the two constraints closer together and try to reverse them to derive the target output. *VOICEDOBS is also (non-crucially) demoted as another winner-preferring constraint.

/bɛt+ən/	IDENT(voi)-IO	*V[-voi]V	*VOICEDOBS
✓ bɛ.d-ən	*!→		**→
☞☹ bɛ.t-ən		←*	*

Tableau 3.8. Output error from the input /bɛt+ən/; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

The same type of output errors also raises both the IEP of singular allomorphs and MEP[X+Ø]$_{SG}$. However, the higher SPs of singular allomorphs, which benefit from the frequency bias, do not decline significantly since output errors are also derived from plural allomorphs: When the input of singular forms is /bɛd/, output errors emerge since the voiced obstruent coda is not repaired with devoicing as shown in Tableau 3.9 repeated from Tableau 3.6 in §4.1. As a result, the IEP of plural allomorphs and MEP[X+ən]$_{PL}$ also increase to lower the SPs of plural allomorphs. Although such errors will gradually promote DEP-IO and MAX-IO over IDENT(voi)-IO, singular allomorphs at this point still have the advantage in the input selection process due to their high token frequency.

/bɛd/	*VOICEDOBSCODA	IDENT(voi)-IO	DEP-IO	MAX-IO
✓bɛt		*!→		
bɛd	*!			
bɛ.di			*!	
☞☹bɛ				←*

Tableau 3.9. Deleting an underlying voiced obstruent coda with the P stage ranking; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

Phase II

Major grammar changes:

{MAX-IO, DEP-IO} » {IDENT(voi)-IO, *V[-voi]V*}

(final devoicing acquired, with optional intervocalic voicing)

Lexical learning:

The SP of plural allomorphs gradually increases after final devoicing has been acquired.

Output patterns:

/bɛt/→[bɛt] (correct output)

/bɛd/→[bɛt] (correct output via final devoicing)

/bɛt+ən/→*[bɛ.t-ən] or [bɛ.d-ən] (optional intervocalic voicing)

/bɛd+ən/→[bɛ.d-ən] (correct output)

/ɑp/→[ɑp] (correct output)

/ɑp+ən/→*[ɑ.b-ən] or [ɑ.p-ən] (optional intervocalic voicing)

The shift from Phase I to Phase II occurs with two temporary changes in the constraint grammar: (i) IDENT(voi)-IO and *V[-voi]V are unranked, and (ii) DEP-IO and MAX-IO strictly dominate IDENT(voi)-IO. The former at times allows the mapping from the input /bɛt+ən/ to the target output [bɛ.d-ən] of APs with an optional intervocalic voicing process; the input /bɛd+ən/ is no longer the only source of the target output. Nevertheless, the side-effect is the intervocalic voicing errors of NAPs derived by the same grammar as in Tableau 3.10 (cf. Tableau 3.8).

/ɑp+ən/	IDENT(voi)-IO	*V[-voi]V
✓ɑ.p-ən		*→
☞☹ɑ.b-ən	←*	

Tableau 3.10. Output errors of non-alternating plural forms caused by optional intervocalic voicing; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

The second change is equal to the acquisition of the correct repair strategy of voiced obstruent coda (i.e. devoicing) as in Tableau 3.11 (cf. Tableau 3.6). With a final devoicing grammar, plural allomorphs do not generate any output errors (i.e. /bɛd/→[bɛt] and /bɛd+ən/→[bɛ-d.ən]). Their IEP and MEP([X+ən]$_{PL}$ now start dropping to gradually raise their SP, and plural allomorphs eventually dominate the input selection process.

/bɛd/	*VOIOBSCODA	DEP-IO	MAX-IO	IDENT(voi)-IO
☞✓bɛt				*
bɛd	*!			
bɛ.di		*!		
bɛ			*!	

Tableau 3.11. Devoicing voiced obstruent coda after promoting DEP-IO and MAX-IO

Phase III:

Major grammar changes:

IDENT(voi)-IO » *V[-voi]V

(converge on the target grammar without intervocalic voicing)

Lexical learning:

Plural allomorphs receive a dominant SP.

Output patterns:

/bɛt/→[bɛt] (correct output)

/bɛd/→[bɛt] (correct output via final devoicing)

/bɛt+ən/→*[bɛ.t-ən] (only when plural allomorphs decay from the lexicon)

/bɛd+ən/→[bɛ.d-ən] (correct output)

/ɑp/→[ɑp] (correct output)
/ɑp+ən/→ [ɑ.p-ən] (optional intervocalic voicing)

Finally in Phase III, after plural allomorphs are assigned a dominant SP, the chance for the input of APs to be /bɛt+ən/ will be lower, and the pressure of moving IDENT(voi)-IO and *V[-voi]V together to derive the target output disappears as well. Since NAPs always require IDENT(voi)-IO to strictly outrank *V[-voi]V, the ranking between the two constraints is thus rebuilt in Phase III. Note that although output errors are expected to be rare after the target grammar is acquired and all alternating plural allomorphs are assigned a high SP, some errors may occur when low frequency plural allomorphs decay from the memory occasionally (see §1.4 of Ch. 2); selecting singular allomorphs as the input like /bɛt+ən/ creates the output error *[bɛ.t-ən] with the ranking IDENT(voi)-IO » *V[-voi]V as in Phase I.

4.3 Grammar learning after morphological awareness

We first examine whether grammar learning in the simulation follows the above predictions. The rapid changes in constraint values of the initial 2,000 learning cycles are presented in Figure 3.1 with the initial values inherited from the P stage simulation in Table 3.6. The development of *VOICEDOBSCODA, ONSET, and *CODA is excluded from Figure 3.1 since the three constraints are not demoted or promoted significantly and are not crucial to the changes occurring in the different phases below.

Before the 500th learning cycle, a few crucial changes occur as expected in §4.2: MAX-IO is slightly promoted along with the demotion of IDENT(voi)-IO for a final devoicing grammar (see Tableau 3.6 and Tableau 3.10 in §4.2), and *V[-voi]V is promoted closer to IDENT(voi)-IO so the target output of APs like [bɛ.d-ən] can be derived from inputs like /bɛt+ən/ with intervocalic voicing (see Tableau 3.8 in §4.2). This interval from the 1st to the 500th learning cycle corresponds to Phase I predicted in the previous section. After the 500th learning cycle, the difference between IDENT(voi)-IO and *V[-voi]V remains insignificant. The emergence of such an optional intervocalic voicing grammar can be mapped to Phase II assumed in the previous section.

	Ranking: *VoiObsCoda» {Ident(voi), Dep} » {Max, Onset} » {*V[-voi]V*, *VoiObs} » *Coda
Value	100 53 52.7 48.5 43.8 40.6 38.3 25.2

Table 3.6. Constraint values/rankings acquired at the end of the P stage simulation

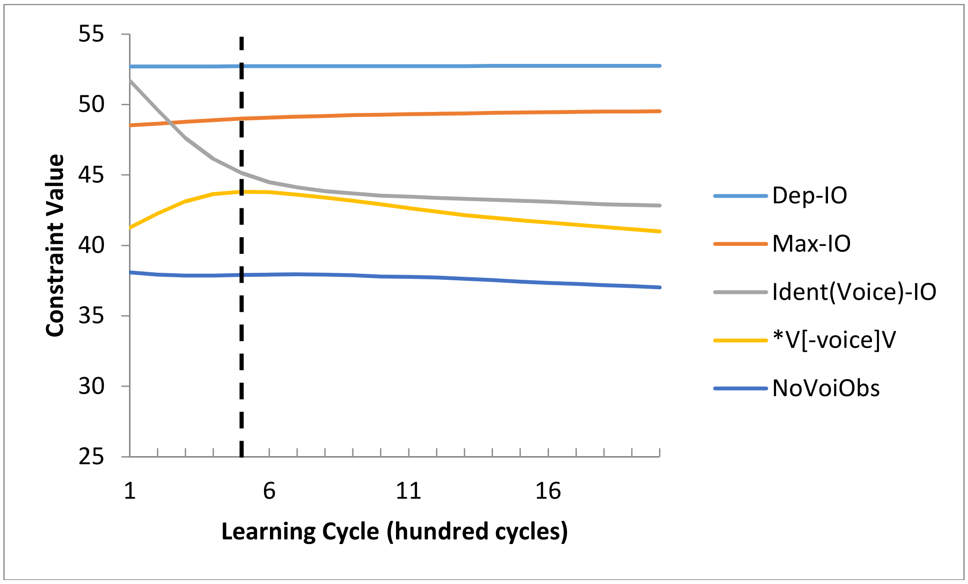


Figure 3.1. Development of constraint values in the initial 2,000 learning cycles in the M stage simulation; the dotted line represents the 500th learning cycle and the boundary between Phase I and Phase II

We can define the developmental stage after the 2,000th learning cycle as Phase III as shown in Figure 3.2, where the further changes of Ident(voi)-IO and *V[-voi]V can be observed. As predicted in Phase III, Ident(voi)-IO regains a more dominant position than *V[-voi]V because singular allomorphs are assigned a lower SP so there is no need to promote *V[-voi]V to derive the target output [bɛ.d-ən] from /bɛt+ən/. Such SP development will be confirmed in the following section.

The final constraint values are listed in Table 3.7, and the translated constraint ranking essentially matches the target ranking in (5) as Dep-IO and Max-IO crucially outrank Ident(voi)-IO which in turn dominates *V[-voi]V. In sum, the algorithm successfully acquires the target grammar and the learning development coincides with the prediction made by PSI-OT-GLA.

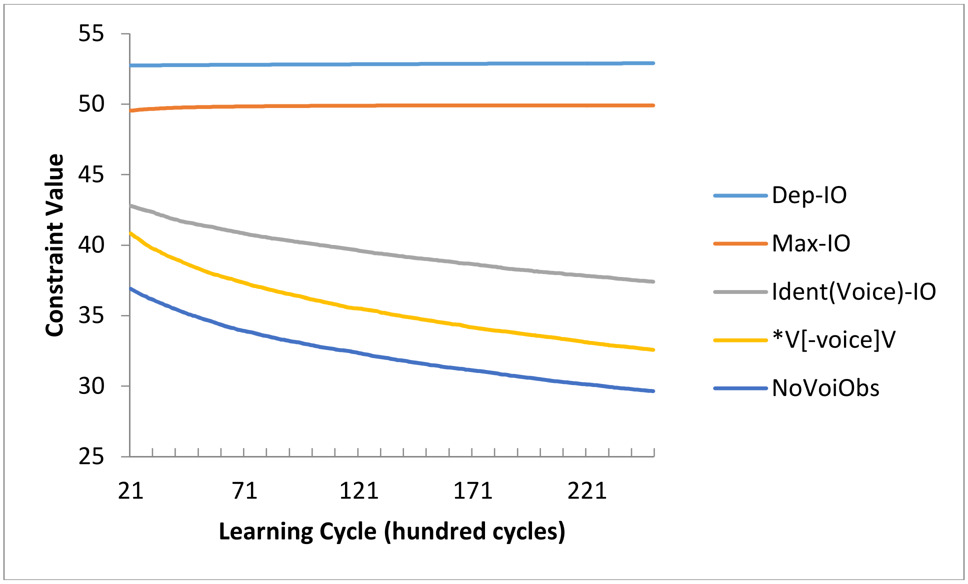


Figure 3.2. Development of constraint values from the 2,100th to the 25,000th learning cycles in the M stage simulation

	Ranking: *VOIOBSCODA» {DEP, MAX} » ONSET » IDENT(voi) » {*CODA, *V[-voi]V*, *VOIOBS}
Value	100 53.1 49.9 43 29.3 25.2 23.8 21

Table 3.7. Constraint values/rankings acquired at the end of the M stage simulation (i.e. the 100,000 cycle)

4.4 Lexical learning after morphological awareness

Now we shift our focus to the assignment of SP to examine whether lexical learning is successful along with grammar learning. The goal of lexical learning is recalled here: Plural allomorphs should be assigned a dominant SP. The SP of each alternating allomorph (‘web’ and ‘crab’ excluded) at the end of the M stage simulation is listed in Table 3.8. The results show that $SP_{\text{‘hand’-PL}}$ and $SP_{\text{‘dog’-PL}}$ are both assigned an absolutely dominant SP of 1. $SP_{\text{‘bed’-PL}}$, while not completely dominant, is still as high as 0.87. Nevertheless, none of $SP_{\text{‘turtle’-PL}}$, $SP_{\text{‘hat’-PL}}$, and $SP_{\text{‘pencil’-PL}}$ is higher than 0.4, allowing the singular allomorph of these stems to be selected as the input much more frequently. Below, I will illustrate the emergence of the predicted learning patterns, explain lower SPs with an extreme frequency bias and the memory decay effect, and discuss whether the results represent a failure in lexical learning.

	bed	turtle	hat	pencil	hand	dog
SP	0.87	0.2	0.38	0.01	1	1

Table 3.8. The SP of plural allomorphs at the end of the M stage simulation

The development of the SP and IEP of the plural allomorph of 'dog' is shown in Figure 3.3 as an example of completely dominant plural allomorphs in the input selection process. Only the development before the 25,000th learning cycle is shown to make initial changes in SP and IEP more visible.

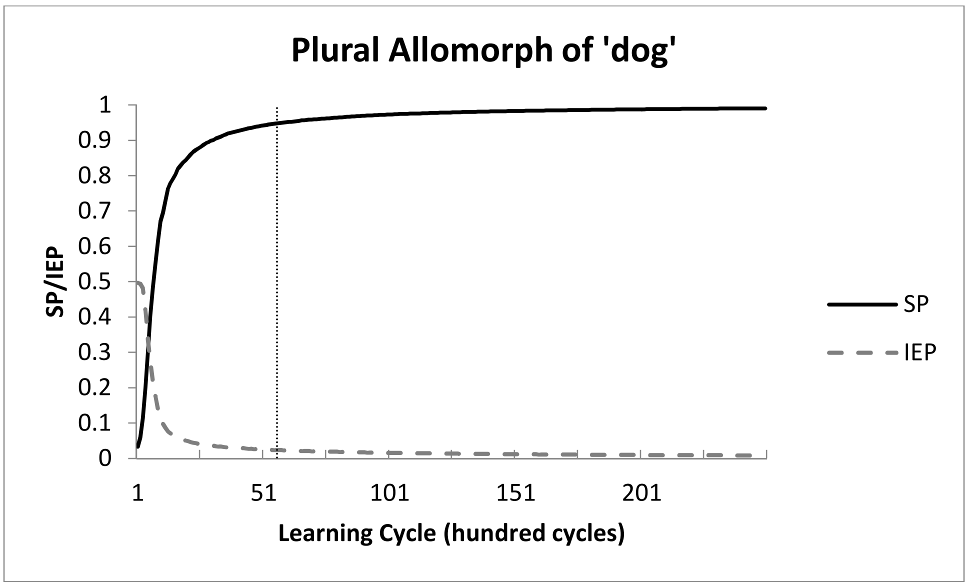


Figure 3.3. Development of the SP (black solid line) and IEP (grey dotted line) of the plural allomorph of 'dog' before the 25,000th learning cycle; the vertical dotted line represents the 2,000th learning cycle (i.e. the boundary between Phase II and III)

Patterns which match the predictions of different phases can be observed in the SP and IEP development. Phase I can be observed with an initially low SP and high IEP in Figure 3.3. The reason of the high IEPs is the dominant IDENT(voi)-IO which prevents a voiced obstruent coda in its singular form from being repaired with devoicing, as shown in §4.3. Deleting the voiced obstruent coda (i.e. /bɛd/→*[bɛ]) or inserting a vowel (i.e. /bɛd/→*[bɛ.di]) raises $IEP_{\text{'dog'-PL}}$. This higher IEP and the frequency bias toward singular allomorphs, result in the initial low SPs in Phase I.

Phase II is demonstrated with the gradually declining $IEP_{\text{'dog'-PL}}$ and the rising $SP_{\text{'dog'-PL}}$ before the 2,000th learning cycle. The changes are triggered by the promotion of MAX-IO over IDENT(voi)-IO, which allows plural allomorphs like /bɛd/ to generate target singular forms via final devoicing (i.e. /bɛd/→[bɛt]). With fewer output errors, $IEP_{\text{'dog'PL}}$ naturally drops and results in a higher $SP_{\text{'dog'-PL}}$, which is in line with the description of Phase II. Finally, the prediction in Phase III states that the plural allomorphs should be assigned a dominant SP, which is true for 'dog' as shown in Figure 3.3.

Unlike the plural allomorph of 'dog' (and 'hand'), the SP of 'bed' only reaches 0.87 by the end of the M stage simulation. The SP/IEP development of the plural allomorph of 'bed' is presented in Figure 3.4 and the difference with Figure 3.3 is apparent: $IEP_{\text{'bed'-PL}}$ does not drop to the floor until the 15,000th learning cycle, and the SP of 'bed' has not reached 0.7 even after 25,000 learning cycles.

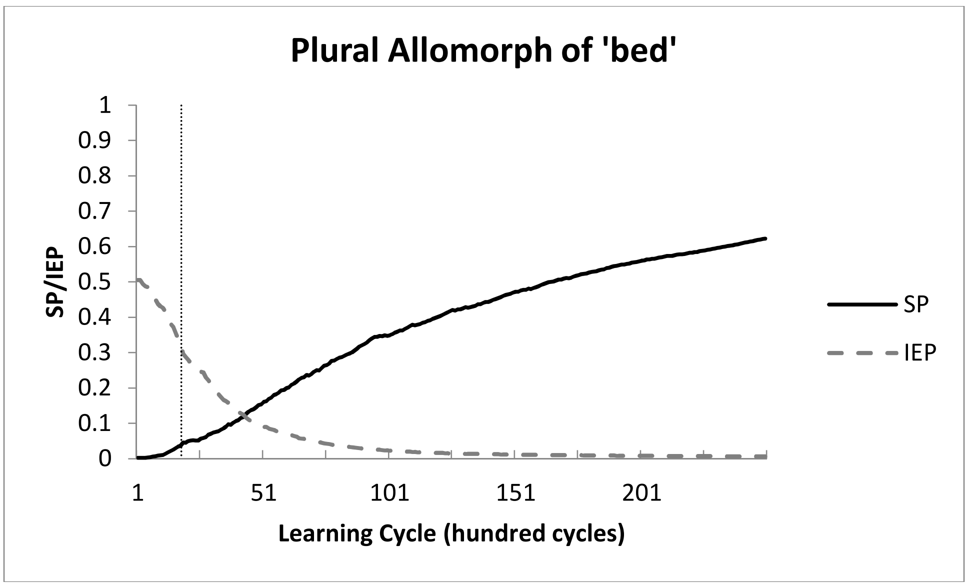


Figure 3.4. Development of the SP (black solid line) and IEP (grey dotted line) of the plural allomorph of 'bed' before the 25,000th learning cycle; the vertical dotted line represents the 2,000th learning cycle (i.e. the boundary between Phase II and III)

The low SP cannot be attributed to the slower decrease of the IEP since the SP is still not dominant when the IEP reaches the floor after the 15,000th cycle. The primary cause of this low SP is the strong frequency bias toward the singular allomorph of 'bed' (23.66:1). $SP_{\text{'dog'-PL}}$ easily increases to almost 1 since the frequency bias toward the singular allomorph is almost negligible (2.02:1, and 1.71:1 for 'hand'). With a frequency ratio of 23.66:1, a low $IEP_{\text{'bed'-PL}}$ is not sufficient to help $SP_{\text{'bed'-PL}}$ counter against the strong frequency bias toward the singular allomorph.

Using the formula introduced in §1.3 of Ch. 2, we can calculate the IEP required for the plural allomorph of ‘bed’ to gain a dominant SP of 0.9 if MEP is not incorporated, as in (8). $IEP_{\text{‘bed’-SG}}$ is 0.036 at the 10,000th learning cycle, and the required $IEP_{\text{‘bed’-PL}}$ must be as low as 0.0015 for an $SP_{\text{‘bed’-PL}}$ of 0.9. That is, the plural allomorph can only produce around 7 output errors out of 1,000 selections. This low IEP is less possible for the plural allomorphs, which can also generate a certain amount of output errors before final devoicing is acquired, as we have seen previously. In the 10,000th learning cycle, $IEP_{\text{‘bed’-PL}}$ is only as low as 0.023, which is still far from the target number calculated in (8).

(8) Predicting the required $IEP_{\text{‘bed’-PL}}$ for an SP of 0.9 with a singular-plural ratio of 23.66:1 without MEP

$$IEP('bed'_{PL}, 0.9) = \frac{(1 - 0.9) \times 1}{0.9 \times (23.66 \times 0.036^{-1})} \approx 0.0015$$

The low token frequency of ‘bed (pl.)’ in the learning input, which is reflected in the learners’ production, also contribute to the slowly growing $SP_{\text{‘bed’-PL}}$. As mentioned in §1.1 of Ch. 2, the speed of lowering IEP of an allomorph is directly correlated with the token frequency of a form in which selecting the allomorph leads to output errors (see the similar effect in the modeling in Jarosz (2011)). In the current case, due to the extremely low token frequency of ‘bed (pl.)’, the input with the singular allomorph (i.e. /bɛt+ən/) is rarely tested, and consequently $IEP_{\text{‘bed’-SG}}$ rises more slowly for learners to recognize the allomorph as a ‘worse’ stem input option.

Both types of the pressure against a dominant $SP_{\text{‘bed’-PL}}$ are nevertheless canceled with the morphological privilege of plural allomorphs (i.e. a higher $MEP[X\text{+ən}]_{PL}$), which aids $SP_{\text{‘bed’-PL}}$ to eventually reaches 0.86 by the end of the M stage simulation. The development of $MEP[X\text{+Ø}]_{SG}$ and $MEP[X\text{+ən}]_{PL}$ from the beginning of the M stage simulation to the 10,000th learning cycle is plotted in Figure 3.5.

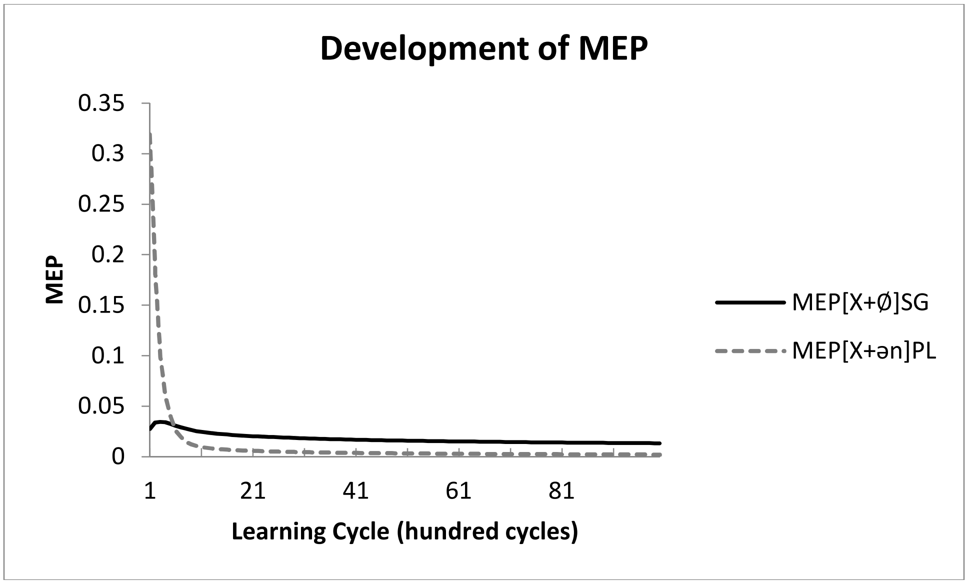


Figure 3.5. Development of the MEP[X+Ø]$_{SG}$ (black solid line) and MEP[X+ən]$_{PL}$ (grey dotted line) from the 1st to the 10,000th learning cycle in the M stage simulation

The reason for an initially higher MEP[X+ən]$_{PL}$ is similar to the higher IEP of each plural allomorphs in Phase I: Output errors emerge as IDENT(voi)-IO dominates MAX-IO and DEP-IO to repair voiced obstruent coda with deletion and insertion, not devoicing. After final devoicing is acquired, plural allomorphs no longer surface as errors, and therefore only MEP[X+ən]$_{PL}$ can freefall to the floor in Phase II. All plural allomorphs eventually benefit from this extremely low MEP[X+ən]$_{PL}$ in the competition against singular allomorphs.

With the help from a low MEP[X+ən]$_{PL}$, the threshold for the plural allomorph of 'bed' to be assigned a dominant SP becomes lower. In (9), IEP$_{\text{'bed'-PL}}$ for an SP$_{\text{'bed'-PL}}$ of 0.9 is calculated with MEP[X+ən]$_{PL}$ (0.002), IEP$_{\text{'bed'-SG}}$ (0.036), and MEP[X+Ø]$_{SG}$ (0.013) in the 10,000th learning cycle. The predicted IEP$_{\text{'bed'-PL}}$ for an SP$_{\text{'bed'-PL}}$ of 0.9 is 0.0099, which is much higher than the threshold calculated in (8). As IEP$_{\text{'bed'-SG}}$ and MEP[X+ən]$_{PL}$ continue to drop to 0.002 and 0.0005 respectively in the 100,000th learning cycle, SP$_{\text{'bed'-PL}}$ is able to conclude at 0.87.

(9) Predicting the required IEP$_{\text{'bed'-PL}}$ for an SP of 0.9 with a singular-plural ratio of 23.66:1 with MEP

$$IEP('bed'_{PL}, 0.9) = \frac{(1-0.9)\times 1}{0.9\times(23.66\times(0.036\times 0.013)^{-1})}\times\frac{1}{0.002}\approx 0.0099$$

The last notable pattern in the lexical learning results is the lower SP of the plural allomorphs of 'pencil', 'hat', and 'turtle' as shown representatively by the development of $SP_{\text{'hat'-PL}}$ and $IEP_{\text{'hat'-PL}}$ in Figure 3.6. These low SPs can be explained with their extremely low token frequency, which makes them more vulnerable to the memory decay effect. When a plural allomorph decays completely from the lexicon at times, its SP must be zero in a given learning cycle; the SP of a low-frequency plural allomorph, when averaged over 100 simulation results, must be lower than the averaged SP of a high-frequency plural allomorph. In addition, since a low-frequency plural allomorph is selected as an input less frequently to surface as a correct output, its number of correct outputs accumulates slowly, which explains the slow decline rate of $IEP_{\text{'hat'-PL}}$.

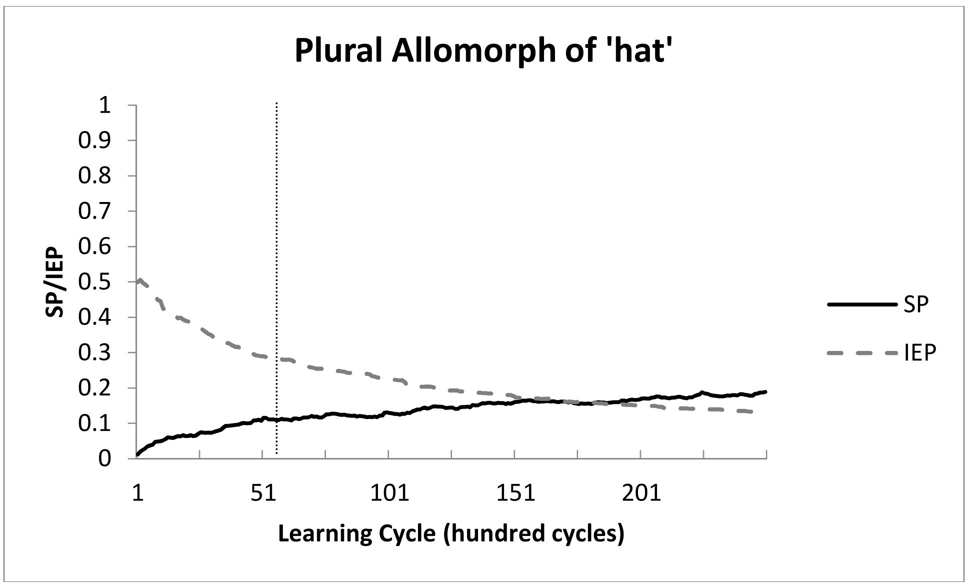


Figure 3.6. Development of the SP (black solid line) and IEP (grey dotted line) of the plural allomorph of 'hat' before the 25,000th learning cycle; the vertical dotted line represents the 2,000th learning cycle (i.e. the boundary between Phase II and III)

To sum up here, PSI-OT-GLA can acquire the crucial constraint ranking *VOICEDOBSCODA » MAX-IO » IDENT(voi)-IO » *V[-voi]V by the end of the M stage simulation, and a dominant SP is correctly assigned to the lexically stable plural alternating allomorphs. Readers may alternatively consider PSI-OT-GLA learning a failure since eventually PSI-OT-GLA is unable to assign a dominant SP to the lexically unstable or low-frequency plural alternating allomorphs. However, note that the end of the current simulation does not necessarily correspond to the real final state in Dutch morphophonological acquisition, and the development may move toward the real final state if the simulation continues. I will return to discuss this point in §4.6 below.

4.5 Simulated results vs. experimental results

The current section is the comparison between the results of the simulated elicitation tasks and the real experimental results summarized in §1. The targets of the comparison are the error rates of plural form production, which are calculated as the number of voicing errors divided by the sum of voicing errors and targets as in §1.

The comparison between the two experiments first requires mapping experiment cycles with different age groups to find presumably similar developmental progress for the machine learner. In the M stage simulation, an elicitation task is simulated for every 50 learning cycles, and the three out of 100,000 / 50 = 2,000 experiment cycles are selected to be compared with the experimental results of three age groups in Kerkhoff (2007). The development of the error rate of plural forms is illustrated in Figure 3.7, which is crucial to map individual learning cycles to age groups.

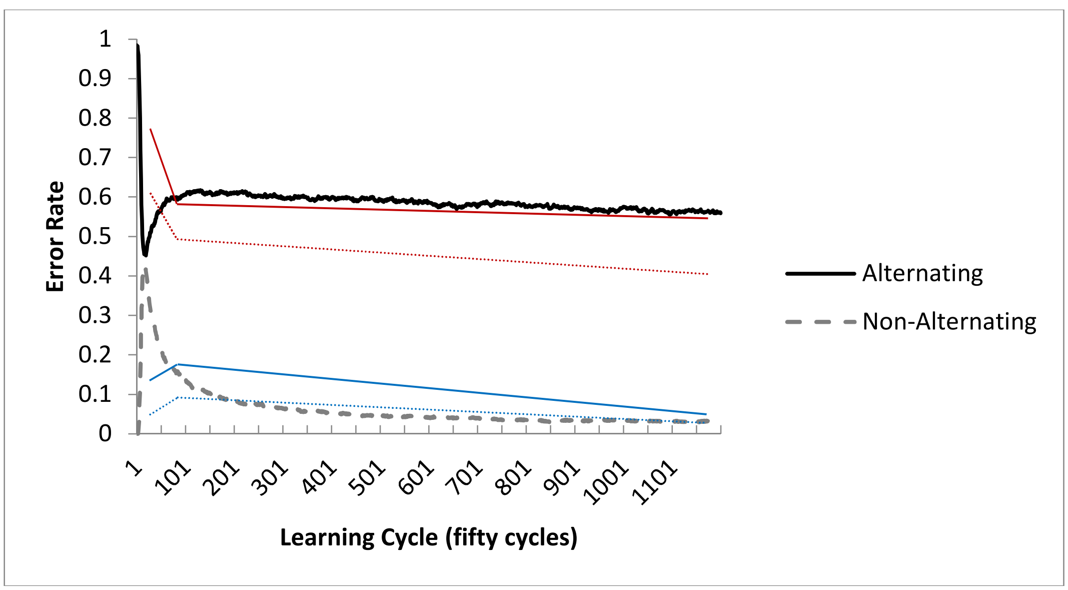


Figure 3.7. Changes in error rates from the 50th learning cycle to the 60,000th learning cycle; color lines represent the trends of alternating plural form (red) and non-alternating plural form (blue) error rate changes between a selected learning cycle (solid) and an age group (dotted)

It can be seen in Figure 3.7 that the AP error rate is extremely high at the very beginning of the M stage simulation, but the NAP error rate is much lower at the same point. This initial stage is consistent with the performance of the youngest age group (2;9-3;11), and the pair of error rates in the 350th learning cycle and those produced by the age group are judged to be comparable as shown in Table 3.9. Later, there is also a point in the very beginning of the M stage simulation in Figure 3.7 where the two error rates are very close to each other, and this pattern occurs due to the frequent application of intervocalic voicing. Comparable error rates are nevertheless not absent in the experimental results, and the results of these learning cycles may not be similar to those of any age group (see discussion in §4.6). The AP error rate later slightly increases again between the 2,500th and 5,000th learning cycle, which is yet still a great improvement from the very beginning. The NAP error rate drops significantly as well but remains higher at the same point than at the onset of the simulation. This development is similar to the error changes from the youngest age group to the second age group (4;0-6;2): The AP error rate decreases with an increase in the NAP error rate. The learning cycle that corresponds to the second age group is thus selected around the 5,000 learning cycle, which is the 3,500th cycle with comparable error rates. Finally, the simulation results shows a mild but continuous improvement in both of the AP and NAP error rate, which is consistent with the improvement from the second to the oldest age group (6;9-7;8) in the experimental results. Again, the 60,000th learning cycle is paired with the oldest age group due to similar error rates. Overall, the three selected learning cycles overestimate the two error rates, and I will return to this issue in §4.6.

	Simulated Experiment			Real Experiment	
Cycle	**AP Error%**	**NAP Error%**	**Age**	**AP Error%**	**NAP Error%**
350th	79.9%	13%	**2;9-3;11**	61%	3.8%
3,500th	59.5%	16.6%	**4;0-6;2**	49.6%	5.4%
60,000th	55.9%	3%	**6;9-7;8**	41.5%	2%

Table 3.9. Comparison of plural form error rates between simulated and real experiments; AP = alternating plural form, NAP = non-alternating plural form

The comparison of alternating plural error rates by item is visualized in Figure 3.8. The error rates from the simulated results are averaged over the three selected learning cycles to be compared with those averaged over the three age groups in Kerkhoff's experiment, and the error rates from the real experimental results are borrowed from §1 with the numbers of devoicing errors and target outputs listed in Table 3.10.

	beds	turtles	hats	pencils	hands	dogs	webs	crabs
error token (Exp)	23	33	32	12	12	9	32	40
target token (Exp)	27	15	12	34	38	41	9	9
error rate (Exp)	46%	69%	73%	26%	24%	18%	78%	82%
error rate (Sim)	60.7%	72.5%	73.4%	84.1%	22.7%	28.1%	87%	87%

Table 3.10. Devoicing error rates in Kerkhoff's experiment and simulation

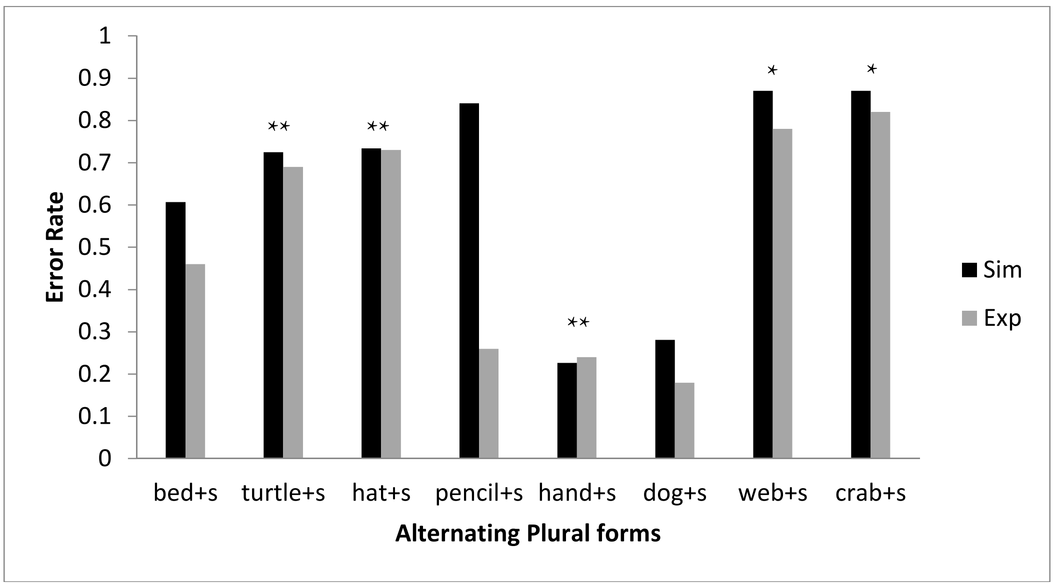


Figure 3.8. Comparison of alternating plural error rates between simulated and experimental results; * difference ≤ 10%, ** difference ≤ 5%

The comparison demonstrates a close approximation between the simulated and real experimental results for six out of eight APs; the difference in the error rate is smaller than 10% for 'turtles', 'hats', 'webs', and 'crabs', and for 'dogs' the difference is 10.1%. The error rate of 'beds' is overestimated in the simulations with a difference of 14.7%. The reason for overestimating the 'beds' errors in the simulation lies in the extreme frequency bias toward the singular allomorph of 'bed' (23.66:1), and the number of output errors generated from the singular allomorph is not high enough to cancel this unbalanced frequency distribution. However, I argue that the similarity between the two error rates of 'beds' stands since if the error rate of 'pencils' is excluded, the error rate of 'beds' is third to the lowest error rate in both the simulation and experimental results.

The primary disparity between the simulated and real experimental results is the error rate of 'pencils', which doubles in the simulated result and is more similar to the error rate of 'turtles' and 'hats'. Recall that since the memory decay rate is 0.002 in the simulation, the plural allomorph of 'pencil', 'turtle', and 'hat' is vulnerable to memory decay. The high error rate of 'pencils' (as well as 'turtles' and 'hats') in the simulated results is thus a product of frequently selecting the singular allomorph as the input and prohibiting intervocalic voicing with the ranking IDENT(voi)-IO » *V[-voi]V. Then, the question is why the error rate is much lower in the experimental results. One possible line is the sampling issue. Although token frequencies in CELEX seem to mirror those in child-directed speech (see §2.1), the input distribution for each Dutch learner may still vary slightly. That is to say, the plural form 'pencils' might be frequently present as learning inputs for some learners but absent for many other learners. The former can still select the plural allomorph of 'pencil' as the input to produce the target output, but the latter cannot. Whatever the reason is (see also §4.6 for some discussion), this single aberration does not vitiate the otherwise substantial similarity in the error rates between the simulated and real experimental results as supported by a significant positive correlation (r = 0.73, p = 0.04). If the obvious outlier 'pencils' is removed, the positive correlation further improves (r = 0.98, $p < 0.001$). The improvements between the two correlation coefficients are approaching significance (z = 1.93, p = 0.054) when compared using Fisher r-to-z transformation (Lowry 2001-2013).

4.6 Discussion

The above modeling successfully accounts for the trend in the error rate changes, but the averaged AP and NAP error rate are clearly overestimated in the simulation. Particularly, it has been shown in Table 3.9 that the averaged error rates of NAPs are never higher than 6% in Kerkhoff's experiment, reflecting the rarity of errors like *[ɑ.b-ən] (the same conclusion was reached with Kerkhoff's corpus study using CHILDES (MacWhinney 1991/2000)). PSI-OT-GLA, however, predicts that the error rate can be as high as 21.8% as in the 3,500th learning cycle – the assumed Phase II during which chance is higher for *V[-voi]V to outrank IDENT(voi)-IO to allow intervocalic voicing.

Despite the above discrepancy, I suggest the simulated and experimental results to be comparable for following reasons. While the average non-alternating error rate in Kerkhoff's experiment is generally very low, there is in fact a great variation across individuals; ten out of 58 subjects produced at least 15% of non-alternating plural errors and two out of ten at the age of 3;4 and 5;11 had an error rate of at least 35% (Kerkhoff 2007:154, Figure 3). In Figure 3.7 above, it was also shown that the averaged NAP error rate could be as high as 40% during the simulation course (and 16% in the selected learning cycle). Therefore, the current series of simulations may in fact predict a more extreme type of learners who frequently produces more overgeneralization patterns as the two subjects in the experiment. Note that Kerkhoff's experiments are not a longitudinal study either. It is thus plausible to speculate that some subjects are in the stage with a higher error rate of NAPs and others, yet the majority of participants, are not at the time of Kerkhoff's study.

There is of course the possibility that the output errors of NAPs are uncommon even in a longitudinal study. If this is the case, PSI-OT-GLA will have to produce these low error rates of NAPs. For this, I propose two possible patches. First, as suggested in Tessier (2006, 2007) and Tessier & Jesney (2011), Output-Output (OO) faithfulness constraints are ranked higher in the very beginning of morphophonological acquisition. Therefore, if IDENT(voi)-OO dominates *V[-voi]V, /ɑp+ən/ will surface as the target [ɑ.p-ən] even if IDENT(voi)-IO is outranked by *V[-voi]V in Tableau 3.12. More details of this alternative model will be discussed in Ch. 6.

/ɑp+ən/ Base: [ɑp]	IDENT(voi)-OO	*V[-voi]V	IDENT(voi)-IO
☞✓ɑ.p-ən		*	
ɑ.b-ən	*!		*

Tableau 3.12. Preserving the intervocalic voicing contrast in Dutch with OO-F

The other remedy is to shorten the time for PSI-OT-GLA to recognize singular allomorphs as 'worse' input options. In the current version, plural allomorphs slowly gain a dominant SP partially because of the frequency bias toward singular allomorphs. Therefore, inputs like /bɛt+ən/ can be constantly selected as the input of APs during this time, and the more the inputs are selected, the more possible IDENT(voi)-IO and *V[-voi]V are pulled together to derive the target output of the APs. Consequently, more output errors like *[ɑ.b-ən] surface due to the grammar shift for the emergence of optional intervocalic voicing. It is thus necessary to somehow converge on the 'better' input option for the AP stems before *V[-voi]V and IDENT(voi)-IO are too close to prevent intervocalic voicing from being extended to NAPs. The implementation of both extensions will be further evaluated, and the current simulated results, although not a perfect match for the real production patterns, should still be considered a close approximation to the actual development of the Dutch morphophonological acquisition.

Another issue in the simulated results is whether PSI-OT-GLA can indeed acquire an adult-like lexical generalization. Particularly, it is expected that real Dutch learners eventually assign a dominant SP to all plural allomorphs, including low-frequency ones, thus producing APs correctly without devoicing errors as Dutch adults do. Therefore, failing to reach this expected 'end state' in Dutch morphophonological acquisition by the end of the M stage simulation, PSI-OT-GLA might not be qualified as an effective learning algorithm in this case. In the current chapter, there has been no such implication that the end of the simulation is mapped to the end state of real morphophonological acquisition. However, this is not saying that there is no commitment in PSI-OT-GLA to capture this desirable end state. It is especially worth noting that although the low-frequency plural allomorphs are not assigned a dominant SP after the final simulation cycle as we have seen in §4.4 above, their SP slowly increases as shown in Figure 3.4 and Figure 3.6. It is therefore highly plausible to assume that their SP can be as dominant if the simulation continues.

The plural allomorph of the stem 'pencil' seems to be an exception to the above developmental trend because based on the settings in the simulation and the low token frequency in the CELEX corpus, it will be constantly subject to memory decay. Therefore, it might be either an inappropriate implementation of a fixed memory decay rate in the simulation or an underestimated token frequency that results in the extremely low SP of the allomorph in the simulation. Only after we sort out these issues, it is possible to judge whether PSI-OT-GLA indeed fails in this particular case. Based on the successful grammar learning after the M stage simulation and the lexical learning toward assigning a dominant SP to plural allomorphs, the adult-like 'end state', while not immediately presenting in the simulation, is fully predicted by PSI-OT-GLA.

5. Testing alternative models with fewer lexical factors

This section serves as a comparison of the performance between PSI-OT-GLA and its variants, which is necessary to validate each lexical parameter incorporated into PSI-OT-GLA. If the same learning results can be generated by excluding a parameter from the model, the parameter is redundant. I will start with a baseline model without incorporating frequency and number of output errors and targets in §5.1. Later, I will add frequency plus IEP without memory decay, then frequency plus MEP without memory decay, and finally PSI-OT-GLA with all lexical factors but memory decay (i.e. token frequency, IEP, and MEP) to create another three different models from §5.2 to §5.4. The frequency-only model in which SP is solely determined by raw token frequency will be ignored here. Previous discussions have shown that the singular allomorph of alternating stems always has a higher token frequency, and thus the frequency-only model can never assign a dominant SP to plural allomorphs with a lower token frequency.

Various elements in the original PSI-OT-GLA remain identical in the four variants. The training corpus and the frequency distribution in the corpus are unchanged. The constraint set, initial ranking, and the demotion bias are also inherited from the original algorithm. Finally, the simulated elicitation tasks are conducted with the same interval (i.e. every fifty learning cycles) and the results are collected with the same calculation of the error rates. In other words, only the contribution from individual lexical variables will be evaluated below. The result from each model will also be averaged over 100 simulations with 50,000 P stage learning cycles and 100,000 M stage learning cycles.

5.1 Naïve Parameter Learning without lexical factors

The first alternative model excludes every lexical factor introduced previously but still implements the probabilistic process of input selection. The question that immediate follows is how to assign and modify the SP of each stored allomorph in this baseline model, and the answer lies in an extension of Yang's (2003) Naïve Parameter Learning (NPL). In NPL, possible specifications of parameters begin with an equal probability; if a parameter is binary, either specification of the parameter has a probability of 0.5 at the very beginning. The probability is adjusted depending on the output; if the output is an error, the probability of the parameter specification used to produce this error is lowered, and if the output is the target, the probability is raised. This approach is essentially similar to how SPs are modified in PSI-OT-

GLA, except that in NPL the amount of each penalty and reward is fixed and not affected by other variables. Here I will extend NPL to the input selection process as a random baseline learner to see whether plural allomorphs can be assigned a dominant SP (henceforth NPL-OT-GLA).[40] If the learning outcome also matches the real experimental results, all the lexical parameters in PSI-OT-GLA may be considered totally unnecessary.

In NPL-OT-GLA, the initial SP of a pair of allomorphs and the adjustments are defined in (10). As in PSI-OT-GLA, if an allomorph is not observed by the learner yet, its SP will be zero until perceiving the first token of the allomorph. When two competing allomorphs have been observed by the learner at least once, the two corresponding SPs are set equally as 0.5 with no reference to their token frequency. For example, the singular allomorph of 'bed' might be perceived 23 times before the first time the plural allomorph of 'bed' is perceived, but at the point where the plural allomorph is perceived, the two SPs are still set as 0.5. When the competition begins, the adjustment amount is fixed at 0.05, which is subtracted from SP with an output error, and is added to the SP with a target output. The upper and lower bound of an SP are still 1 and 0 respectively; if adding or subtracting 0.05 results in a number higher than 1 or lower than 0, the change will not be made.

(10) Initial SPs of a pair of singular and plural allomorph and the adjustments

When learning cycle = 1,

$SP_{SG} = 0.5$

$SP_{PL} = 0.5$

If output ≠ target & input = SG & $(SP_{SG} - 0.05) \geq 0$

$SP_{SG} = SP_{SG} - 0.05$

Else if output ≠ target & input = PL & $(SP_{PL} - 0.05) \geq 0$

$SP_{PL} = SP_{PL} - 0.05$

Else if output = target & input = SG & $(SP_{SG} + 0.05) \leq 1$

$SP_{SG} = SP_{SG} + 0.05$

Else if output = target & input = PL & $(SP_{PL} + 0.05) \leq 1$

$SP_{PL} = SP_{PL} + 0.05$

[40] See also Jarosz (to appear) for a variant of OT based on NPL (Naïve Pairwise Ranking Learner) to solve the hidden structure problem.

Table 3.11 shows that the crucial rankings such as {DEP-IO, MAX-IO} » IDENT(voi)-IO and IDENT(voi)-IO » *V[-voi]V are acquired in NPL-OT-GLA, which is the key to assign a higher SP to plural allomorphs. After 100,000 learning cycles in the M stage simulation, $SP_{\text{'dog'-PL}}$, $SP_{\text{'hand'-PL}}$, and $SP_{\text{'bed'-PL}}$ have an SP higher than 0.9 as shown in Table 3.12, which is a pattern demonstrated in PSI-OT-GLA. The other three SPs (i.e. $SP_{\text{'pencil'-PL}}$, $SP_{\text{'hat'-PL}}$, and $SP_{\text{'turtle'-PL}}$), however, are also ranged from 0.69 and 0.79, which significantly differ from the learning results in PSI-OT-GLA. We will see the impact of this distinction below.

	Ranking: *VOIOBSCODA» {DEP, MAX} » ONSET » *CODA »IDENT(voi) » {*V[-voi]V*, *VOIOBS}
Value	100 53.1 50.5 43 24.9 14.7 9.6 6.8

Table 3.11. Constraint values/rankings acquired at the end of the M stage simulation in NPL-OT-GLA

	bed	turtle	hat	pencil	hand	dog
SP	0.91	0.7	0.79	0.69	1	0.99

Table 3.12. The SP of plural allomorphs at the end of the M stage simulation in NPL-OT-GLA

As in §3.5, three learning cycles corresponding to the three age groups are selected following the same criteria such as similar error rates and an identical tendency in the rate changes across different selected cycles or age groups, which are represented in Figure 3.9 and Table 3.13. Unlike PSI-OT-GLA, the error rates of APs in the selected learning cycles are more similar to those in the real experiments, but the error rates of NAPs are still slightly overestimated in NPL-OT-GLA.

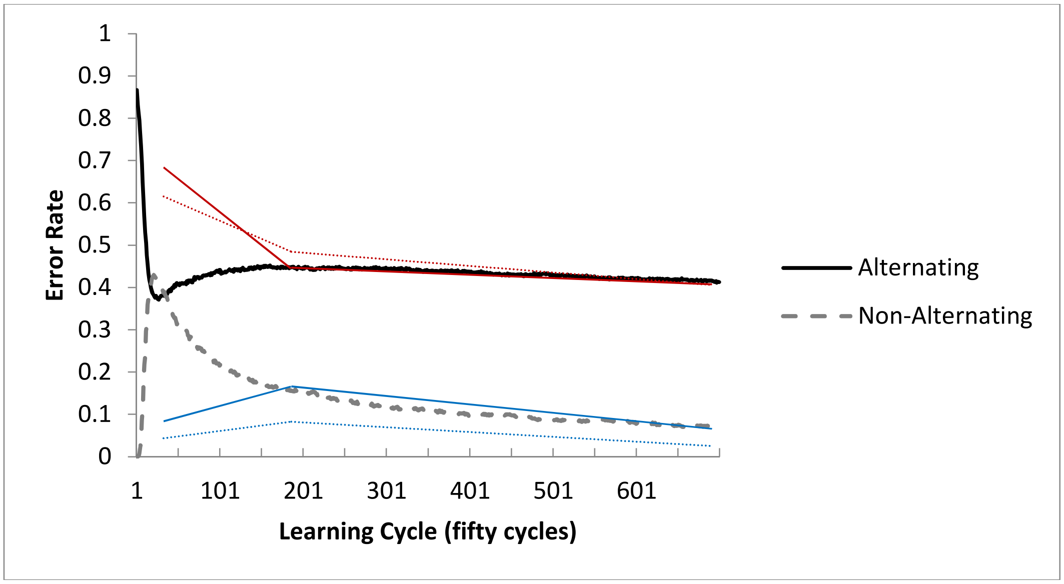


Figure 3.9. Changes in error rates from the 50th learning cycle to the 35,000th learning cycle; color lines represent the trends of AP (red) and non-alternating plural form (blue) error rate changes between a selected learning cycle (solid) and an age group (dotted)

	NPL-OT-GLA		Real Experiment		
Cycle	AP Error%	NAP Error%	Age	AP Error%	NAP Error%
350th	69.1%	8.6%	2;9-3;11	61%	3.8%
8,000th	44.9%	16.5%	4;0-6;2	49.6%	5.4%
35,000th	41.2%	7.3%	6;9-7;8	41.5%	2%

Table 3.13. Comparison of plural output error rates between simulated and real experiments

Visual inspection of Figure 3.10 immediately reveals some major differences between the results produced by PSI-OT-GLA and NPL-OT-GLA. While producing a similar error rate for 'beds', NPL-OT-GLA greatly underestimates the error rate of the APs 'turtles', 'hats', and 'hands', although for 'pencils' such underestimation generates an error rate similar to the one in the experimental results. The difference in the error rate of the four plural forms is between 16.6% (turtles) and 25% (hats). Thus, if 'pencils' is also excluded from the correlation test as in §3.5, the positive correlation is weaker ($r = 0.87$, $p = 0.01$). If this positive correlation is compared with the correlation between the PSI-OT-GLA results and the experimental results ($r = 0.98$, $p < 0.001$) using the Fisher r-to-z transformation, the latter is somewhat higher but does not reach the significance level ($z = 1.24$, $p = 0.22$). All in all, the simulated results suggest that NPL-OT-GLA does not perform better than PSI-OT-GLA, despite the fact that it ended up acquiring the target Dutch constraint ranking and the expected lexical generalization faster than PSI-OT-GLA. The primary reason is that NPL-OT-GLA considerably underestimates the number of output errors for some plural forms as discussed above.

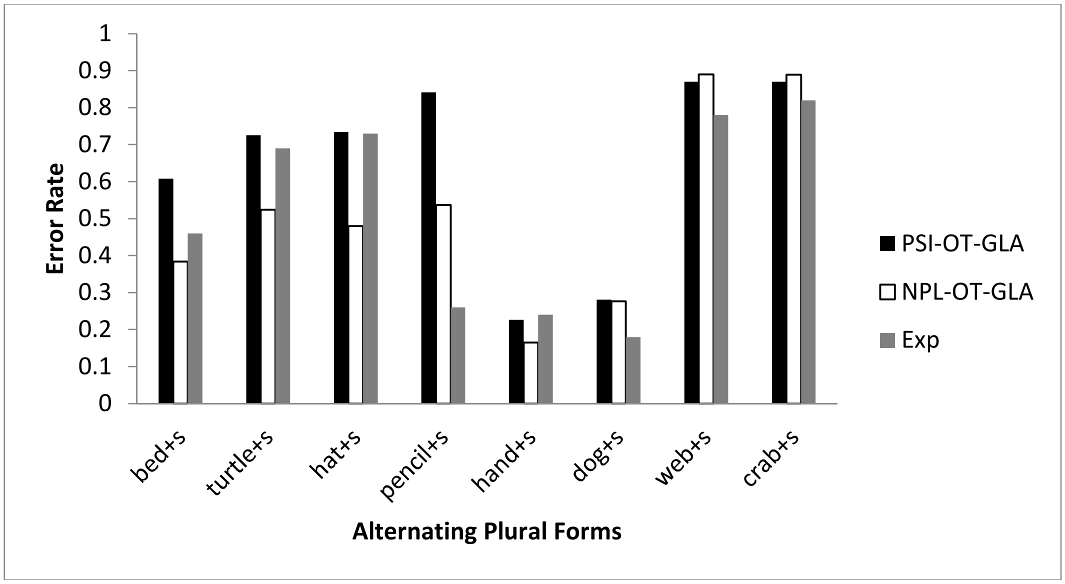


Figure 3.10. Comparison of alternating plural error rates between NPL-OT-GLA, PSI-OT-GLA, and experimental results

5.2 Frequency plus Individual Error Proportion

The second alternative model is the original PSI-OT-GLA excluding MEP and memory decay. That is, only the raw token frequency and IEP of each allomorph contribute to the SP of the allomorph (see §1.2 of Ch. 2).

From the result of grammar learning in Table 3.14, we can see that the model fails to acquire the crucial ranking IDENT(voi)-IO » *V[-voi]V. The force that pulls the two constraints together has been explained previously: When singular allomorphs are selected as the stem input of APs, *V[-voi]V needs to outrank IDENT(voi)-IO to derive target outputs. Therefore, failing to acquire this crucial ranking implies that the model is unable to assign a dominant SP to plural allomorphs. As predicted, by the end of the M stage simulation, only $SP_{\text{'hand'-PL}}$ is higher than 0.5 (i.e. 0.59), as shown in Table 3.15, and overall singular allomorphs are selected as the stem inputs more often.

	Ranking: *VoiObsCoda» {Dep, Max} » Onset » *Coda »{*V[-voi]V*, Ident(voi), *VoiObs}
Value	100 53 49.6 42.9 25.1 -5.5 -5.5 -7

Table 3.14. Constraint values/rankings acquired at the end of the M stage simulation in the Freq+IEP model

	bed	turtle	hat	pencil	hand	dog
SP	0.05	0.49	0.15	0.22	0.59	0.44

Table 3.15. The SP of plural allomorphs at the end of the M stage simulation in the Freq+IEP model

In §1.2 of Ch. 2, I pointed out two reasons why IEP alone cannot assign a dominant SP to plural allomorphs. First, the allomorphs from the same morphosyntactic context cannot benefit from each other with positive results produced by any member of these allomorphs. Second, IEP alone is never sufficiently low for a completely dominant SP. For example, $IEP_{\text{'dog'-PL}}$ is 0.16 at the end of the M stage simulation. However, with an $IEP_{\text{'dog'-SG}}$ of 0.39 at the same point and a singular-plural ratio of 2.02:1, $IEP_{\text{'dog'-PL}}$ must be as low as 0.02 for a dominant $SP_{\text{'dog'-PL}}$ of 0.9, as shown in (11).

(11) Predicting the required $IEP_{\text{'dog'-PL}}$ for an SP of 0.9 with a singular-plural ratio of 2.02:1 without MEP

$$IEP('bed'_{PL}, 0.9) = \frac{(1-0.9)\times 1}{0.9\times(2.02\times 0.39^{-1})} \approx 0.02$$

This model is therefore ruled out as an alternative to PSI-OT-GLA by failing to acquire the target grammar and assign a dominant SP to at least some plural allomorphs, regardless of the error rates of APs produced by this model.

5.3 Frequency plus Morphological Error Proportion

Another possible variant is to remove only IEP from the original PSI-OT-GLA model, so SP is only determined by token frequency and MEP. This model has the potential to succeed since alternating plural allomorphs should in theory be preferred with the global effect of MEP, despite the lack of individual variations caused by IEP. The original formula of Adjustment Token Frequency (ATF) is thus revised as (12).

(12) The Adjusted Token Frequency (ATF) of some allomorph A is defined as follows, where *freq(A)* is the raw token frequency of some allomorph A and *MEP(C)* is the Morphological Error Proportion of the context C where A is from.

$$ATF(A) \equiv freq(A) \times MEP(C)^{-1}$$

The outcome of grammar learning in Table 3.16, nevertheless, sharply contrasts with the above prediction of a successful model. Like the frequency plus IEP model, this variant fails to acquire the crucial ranking IDENT(voi)-IO » *V[-voi]V, which is because the model is incapable of assigning a dominant SP to plural allomorphs. As shown in Table 3.17, none of the plural allomorphs SPs is higher than 0.5, and singular allomorphs are allowed to be selected as inputs more frequently. The pressure to move IDENT(voi)-IO and *V[-voi]V is thus created when the target output [ɑ.b-ən] must be derived from the input /ɑp+ən/ via intervocalic voicing.

	Ranking: *VOIOBSCODA» {DEP, MAX} » ONSET » *CODA »{*V[-voi]V*, IDENT(voi), *VOIOBS}
Value	100 53 49.4 43 25.2 -5.6 -5.7 -7.1

Table 3.16. Constraint values/rankings acquired at the end of the M stage simulation in the Freq+MEP model

	bed	turtle	hat	pencil	hand	dog
SP	0.07	0.46	0.18	0.26	0.5	0.41

Table 3.17. The SP of plural allomorphs at the end of the M stage simulation in the Freq+MEP model

In previous discussion, it has been shown that the plural allomorphs should demonstrate a global tendency of generating fewer output errors, and therefore a low $MEP[X+a]_{PL}$. However, the single lexical parameter, either IEP or MEP alone, is not sufficiently low to produce a dominant SP of plural allomorphs. In the last learning cycle of the M stage simulation, $MEP[X+a]_{PL}$ is 0.02. Nevertheless, the calculation in (13) below indicates that with an $MEP[X+\text{Ø}]_{SG}$ of 0.04 in the same cycle, $MEP[X+a]_{PL}$ should be as low as 0.002 for an SP of 0.9 even for the lowest singular-plural ratio is the 1.71:1 of the stem 'hand'. The Freq+MEP model is thus marginalized with the failure of learning the target grammar and assigning a dominant SP to plural allomorphs.

(13) Predicting the required MEP[X+a]$_{PL}$ for an SP$_{\text{'hand'-PL}}$ of 0.9 with a singular-plural ratio of 1.71:1 without IEP

$$MEP([X+a]_{PL}, 0.9) = \frac{(1-0.9)\times 1}{0.9\times(1.71\times 0.04^{-1})} \approx 0.002$$

5.4 PSI-OT-GLA without memory decay

After the simulation in §5.2 and §5.3, it should be clear that neither IEP nor MEP is dispensable. The issue on the table now is how much the involvement of the memory decay effect can bring the simulated results closer to the real experimental results. In §2.1, I have set the memory decay rate as 0.002 per learning cycle, which forces some low-frequency allomorphs to fade completely from the lexicon at times. In this variant, I will set the memory decay rate to zero and see whether the results – the error rates of the low-frequency plural forms in particular – change significantly.

Grammar learning in this no-decay model is similar to PSI-OT-GLA as shown in Table 3.18; all the crucial rankings are acquired, including the ranking IDENT(voi)-IO » *V[-voi]V that some above alternatives cannot capture. This outcome is not surprising because the removal of memory decay increases the stability of lexical storage, and singular allomorphs are not forced to be the stem inputs since plural allomorphs are decayed from the lexicon. The pressure which requires the grammar to derive the target output via intervocalic voicing is thus weakened.

	Ranking: *VOIOBSCODA» {DEP, MAX} » ONSET » IDENT(voi) » {*V[-voi]V*, *CODA, *VOIOBS}
Value	100 53 50 42.9 31.5 25.6 25.2 22.9

Table 3.18. Constraint values/rankings acquired at the end of the M stage simulation in the PSI-OT-GLA without memory decay

Without the pressure of memory decay, all plural allomorphs in Table 3.19 are assigned an absolutely dominant SP, which is the most noticeable difference between this alternative and PSI-OT-GLA. In particular, the low-frequency plural allomorph of 'turtle', 'hat', and 'pencil' is no longer 'forgotten' to occasionally result in a zero SP, which lowers the SP averaged across 100 simulations.

	bed	turtle	hat	pencil	hand	dog
SP	0.97	1	0.99	1	1	1

Table 3.19. The SP of plural allomorphs at the end of the M stage simulation in PSI-OT-GLA without memory decay

The developmental tendency of the error rates is illustrated in Figure 3.11 showing the three selected learning cycles corresponding to the three age groups in the real experiment, which are the 350th, 2,650th, and 6,000th cycles as in Table 3.20. The error rate of alternating plural allomorphs is only significantly overestimated in the 350th learning cycle, but as in PSI-OT-GLA and NPL-OT-GLA, the error rates of non-alternating plural allomorphs are overall overestimated. The development tendency of both types of error rates is reminiscent of the path in the real experiment. The much higher NAP error rate in the 6,000th learning cycle than in the experimental results does not represent a generally worse performance of this model; the 6,000th learning cycle is selected to align with the oldest age group simply because there is no comparable AP error rate after this cycle. In fact, the NAP error rate in the 6,000th learning cycle in this model is lower than in PSI-OT-GLA (see Figure 3.7 in §4.5).

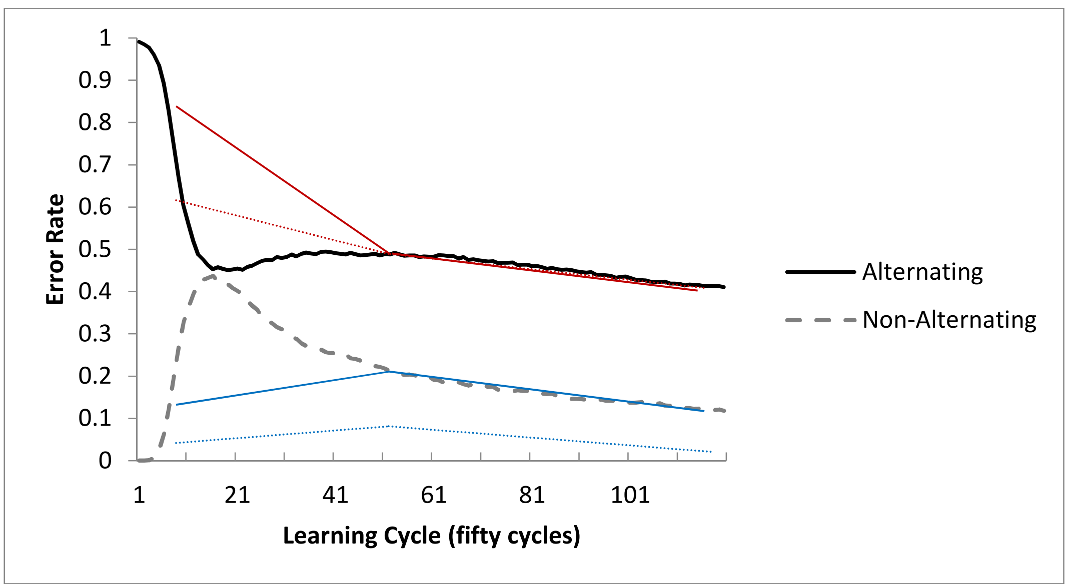


Figure 3.11. Changes in error rates from the 50th learning cycle to the 6,000th learning cycle; color lines represent the trends of AP (red) and non-alternating plural form (blue) error rate changes between a selected learning cycle (solid) and an age group (dotted)

	PSI-OT-GLA (no decay)		Real Experiment		
Cycle	**AP Error%**	**NAP Error%**	**Age**	**AP Error%**	**NAP Error%**
350th	82.8%	12.4%	**2;9-3;11**	61%	3.8%
2,650th	49.2%	20.9%	**4;0-6;2**	49.6%	5.4%
6,000th	41.1%	11.8%	**6;9-7;8**	41.5%	2%

Table 3.20. Comparison of plural output error rates between simulated and real experiments; AP = alternating plural form, NAP = non-alternating plural form

Despite the successful grammar and lexical learning, the by-item comparison in Figure 3.12 still differs from the simulated results of the original PSI-OT-GLA in various ways. Two major distinctions are the underestimated error rate of 'hats' and 'turtles' and the overestimated error rate of 'beds' and 'dogs' in the current result, and the former is closely related to the removal of memory decay. For the stems 'turtle' and 'hat', the frequency bias toward their singular allomorph is weaker (2:1 and 7.75:1), which can be easily cancelled with a low IEP and MEP when there is no memory decay. The overestimated error rate of 'beds' is the result of the slow increase of $SP_{\text{'bed'-PL}}$ if compared to $SP_{\text{'turtle'-PL}}$ and $SP_{\text{'hat'-PL}}$ due to a strong frequency bias toward the singular allomorph of 'bed' (23.66:1). For 'hands', 'webs', and 'crabs', the absence of the memory decay effect does not lead to any obvious difference between the results of the two models. For 'pencils', the model without memory decay greatly lowers the number of output errors but still overestimates the error rates by 32%. The same correlation test excluding 'pencils' is also significantly positive (r = 0.79, p = .04), but the correlation is stronger in the original PSI-OT-GLA (trending; z = 1.64, p = 0.1).

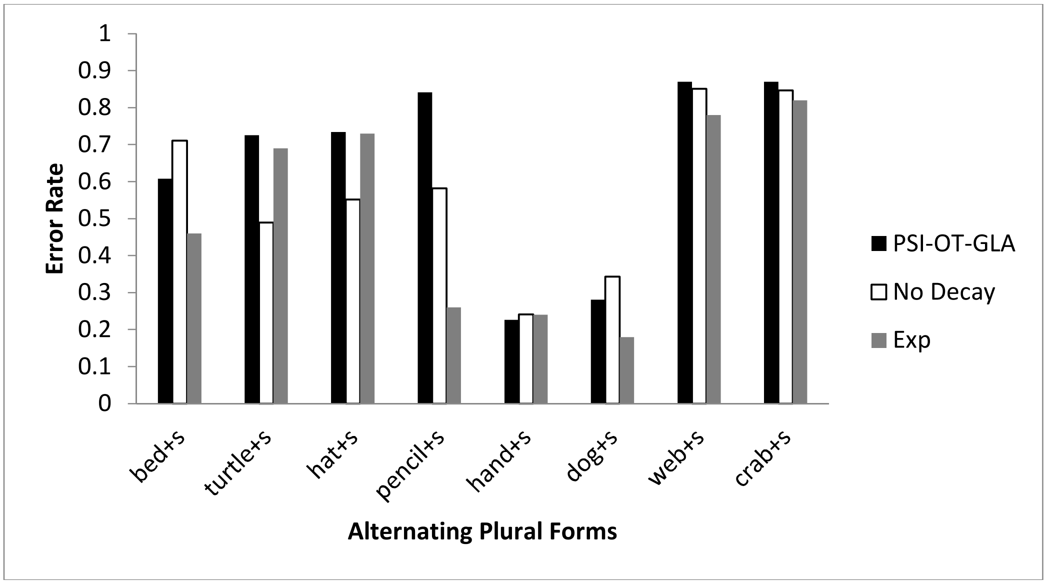


Figure 3.12. Comparison of alternating plural error rates between PSI-OT-GLA, PSI-OT-GLA with memory decay, and experimental results

5.5 Summary of model comparison

The above simulations based on variants of PSI-OT-GLA generated different results. In a baseline model without any lexical factors or a model (§5.1) or a no-decay model (§5.4), the target grammar can be learned correctly. In fact, both models converge on the adult-like lexical generalization (i.e. a dominant SP for every plural allomorph) faster than PSI-OT-GLA. This result is natural since the two primary factors that prevent learners from assigning a dominant SP to plural allomorphs (i.e. frequency bias toward singular allomorphs and memory decay) are absent in the two alternatives. However, these factors are also indispensable in order to account for the word-specific output error patterns, and by discarding them, the two alternatives make less precise word-specific predictions. The two alternatives therefore do not supersede PSI-OT-GLA simply because of a preferable higher efficiency in converging on the expected ‘end state’ (see §4.6 for the discussion on how PSI-OT-GLA may eventually converge on this ‘end state’).

Without IEP (§5.2) or MEP (§5.3), a model cannot even acquire the target grammar. From these comparisons, it is demonstrated that the results generated by PSI-OT-GLA best approximate the real learners’ performance. Furthermore, all of the lexical factors are shown to be crucial to a successful model of learning Dutch stem-final voicing alternations.

6. Local summary

In this chapter, I have demonstrated how PSI-OT-GLA can generate a learning outcome similar to the experiment performance of real Dutch learners and correctly acquire the target constraint grammar. In particular, the majority of by-item error rates of plural forms at different stages are replicated in the simulated results. These errors can be attributed to frequency biases that allow the selection of singular allomorphs of alternating noun stems as the input of APs and a following temporary grammar shift toward intervocalic voicing. Various models were also modified from PSI-OT-GLA by excluding some variables, and none of them could produce a comparable outcome in terms of precisely predicting the performance in various intermediate stages, which in turn highlighted the explanatory power of PSI-OT-GLA. That is, through the simulated results that are similar to real learning performance, we gain a further understanding of morphophonological acquisition. Specifically, we learn that learners have to compute over a set of surface-true allomorphs and some related lexical factors stored in a rich lexicon to probabilistically determine basic allomorphs at least as part of their morphophonological learning strategy. By incorporating this rich lexicon, we can account for not only the acquisition of morphophonemic alternations as a whole (e.g. how plural allomorphs are considered more basic and how a final devoicing grammar is acquired) but also word-specific patterns (e.g. how these developmental stages vary across different forms of different stems). In sum, the possibility of the rich lexicon is first supported by various studies reviewed in Ch. 1, and its necessity in morphophonological acquisition is further endorsed by simulated results comparable to real learning performance discussed in this chapter.

Chapter 4

Probabilistic Selection of Input and Tone 3 change in Standard Mandarin

In the previous chapter I have demonstrated how PSI-OT-GLA can successfully model the acquisition of Dutch morphophology with final devoicing, particularly the by-item error rates and the intermediate stage where overgeneralization occurs as a result of selecting different allomorphs as the input of a morpheme. Although under many circumstances these intermediate stages of overgeneralization are temporary, some may be carried into adulthood when learners continue to fail to select the allomorphs that can surface as target outputs across different contexts more efficiently (e.g. low Individual Error Proportion); instead, learners simply prefer the allomorphs that might have a much higher token frequency.

This chapter focuses on such a case of persistent overgeneralization: Tone 3 in Standard Mandarin, which is undergoing a diachronic change from a concave tone to its shortened low-falling surface variant. I propose a PSI account that the low-falling surface variant, instead of the full concave variant, is selected more frequently as the input of Tone 3 due to a much higher token frequency. The low-falling input then surfaces faithfully even in phrase-final positions, which triggers a gradual grammar change toward banning a concave tone across all prosodic contexts.

In §1, the standard phonological analysis of tone sandhi patterns in Standard Mandarin is reviewed. §2 serves as a summary of the phonetic and phonological differences in lexical tones between two variants of Standard Mandarin – Beijing and Taiwan Mandarin, particularly the observation that the concave tone has already been shortened in the Taiwan variant. A PSI-based hypothesis is proposed to explain this diachronic development. Elicitation tasks were designed to test the hypothesis with the production of Tone 3 words in Standard Mandarin and the results are discussed in §3. Finally, the diachronic development is modeled with computer simulation under the PSI-OT-GLA framework in §4.

1. Tone Sandhi in Standard Mandarin

Standard Mandarin is well-known for its tonal system, which consists of four lexical tones transcribed in Chao's (1930) 5-digit system in (1): 55 (high level), 24 (rising), 213 (concave), and 51 (falling). The phonetic realization of each lexical tone is illustrated in Figure 4.1.

(1) Four lexical tone in Standard Mandarin

Tone 1 high level	ma^{55}	'mother'
Tone 2 rising	ma^{24}	'hemp'
Tone 3 concave	ma^{213}	'horse'
Tone 4 falling	ma^{51}	'scold'

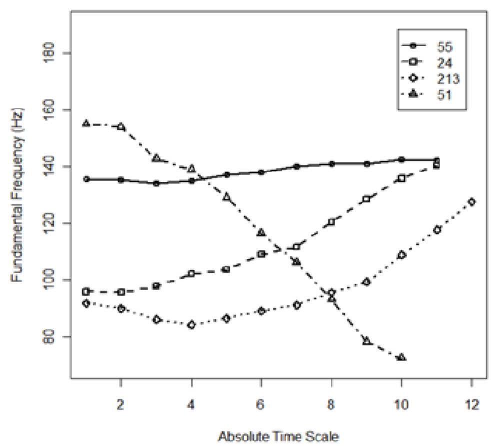


Figure 4.1. Production of [ma^{55}], [ma^{24}], [ma^{213}], and [ma^{51}] by a male native speaker of Standard Mandarin; tonal duration is normalized as a 12-point scale by setting the longest tone (Tone 3, 213) as the reference level

Among the four lexical tones, the concave tone 213, or Tone 3, is involved in two major obligatory tone sandhi rules listed in (2) (e.g. Chen 2000 and Duanmu 2007): Tone 3 Sandhi (T3S) in (2a) only applies when a Tone 3 is followed by another Tone 3, forcing the former to surface as the rising tone 24, whereas Half Sandhi (HS) in (2b) applies whenever a Tone 3 is in a non-final position (i.e. immediately followed by any tone T), reducing the full concave tone to a 'half' dipping tone.[41]

[41] Note that T3S applies with no regard for whether Tone 3 in the target or environment is reduced by HS as we will see in the next section. Therefore, the ordering of the two rules seems arbitrary.

There are other optional tone sandhi processes in Standard Mandarin (e.g. Hyman 1975) but T3S and HS are two primary tonal rules related to the diachronic change of Tone 3 in Standard Mandarin as I will illustrate in §2.

(2) Two major tone sandhi rules in Standard Mandarin

a. Tone 3 Sandhi (T3S): 21(3)→24 / ____ 21(3)

   tɑ²¹³ kou²¹³ → tɑ²⁴ kou²¹³ 'beat dog'

b. Half Sandhi (HS): 213→21 / ____ T

   tɑ²¹³ tʰiŋ⁵⁵ → tɑ²¹ tʰiŋ⁵⁵ 'enquire'

   wu²¹³ ʂɤ²⁴ → wu²¹ ʂɤ²⁴ 'fifty'

   bɑŋ²¹³ tɕiɑ⁵¹ → bɑŋ²¹ tɕiɑ⁵¹ 'kidnap'

The traditional generative view on T3S is a synchronic process triggered by Obligatory Contour Principle (OCP; Leben 1973) which forbids two adjacent identical segments, assuming that contour tones are not composed of individual level tones as in African tone languages but single tonal units (e.g. Bao 1999 and Yip 1980; cf. Chen 2010 and Duanmu 1994). Alternatively, T3S may simply be a historical residue of a diachronic tonal change from Chinese in the 16th century. In Mei's (1977) reconstruction, Tone 3 was a low-level tone 22, and Tone 2 was a low-rising tone 13, and thus the rule (2a) can be rewritten as (3), which is more reminiscent of a typical African tone sandhi rule. Since T3S is applied without any exception synchronically, I will consider it a phonological pattern that needs to be captured in the formal constraint-based analysis below.[42]

(3) Tone 3 Sandhi in the 16th century

22→13 / ____ 22

HS has a clear phonetic motivation in terms of the ease of articulation. Following Zhang (2002, 2004, et seq.), Zhang & Lai (2010) claim that HS is triggered

---

[42] Zhang & Lai (2010) argue that after the tonal change from 22 to 213, T3S is less phonologically natural due to the lack of either articulatory or perceptual motivation and is thus lexicalized (cf. Chen 2008, 2010). In a series of wug tests (see also Zhang et al. 2011), Zhang & Lai found while T3S unexceptionally applied to all novel disyllabic sequences and changed the first concave contour to a rising contour, the rising output of T3S in novel words was not completely neutralized with the rising output in real lexical items. They thus concluded that T3S, which was once a fully general phonological rule, is now lexicalized (cf. Chen 2008, 2010).

by a mismatch between the shorter phonetic duration in a non-phrase-final position and the longer tonal duration required to complete a concave contour. That is, when Tone 3 is shortened from a full concave contour to a simple low dipping contour in a non-phrase-final position, speakers can avoid producing too many tonal ups and downs with shorter phonetic duration. This restriction is a cross-linguistic tendency supported by Zhang's (2002) survey: Phonetically longer contour tones are forbidden in a phonetically shorter non-final position regardless of the typological category of tone languages (i.e. African, Asian, Athabaskan, etc.).

For the upcoming constraint-based simulation in PSI-OT-GLA, a Standard OT analysis is provided in this chapter, starting with the definition of the tonal representation of Tone 2 and Tone 3 in Figure 4.2. The analysis here does not attempt to solve the debate over whether contour tones are units in Standard Mandarin (see Chen 2010, 2013; Zhang 2014), and Tone 2 and Tone 3 are simply assumed to be contour tone units with different feature specifications. The simple combinations of [high] and [low] are equal to three level pitches H ([+high, -low]; corresponding to 5 and 4 in a 5-digit system), M ([-high, -low]; corresponding to 3), and L ([-high, +low]; corresponding to 2 and 1). The representation of Tone 2 thus corresponds to its rising pitch contour MH, and the representation of Tone 3 is the corresponding linear falling-rising sequence MLH of a concave tone.

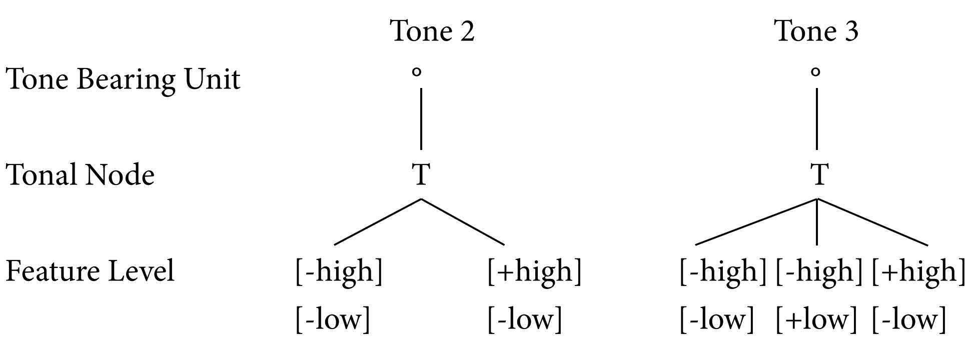


Figure 4.2. Tonal representation of Tone 2 and Tone 3 in Standard Mandarin

The assumption of contour tone units means the adoption of the generative view on the motivation for T3S – i.e. the violation of Obligatory Contour Principle (OCP; Leben 1973) with two adjacent identical contour tone units. Therefore, the top-ranked markedness constraint OCP-MLH (3a) is proposed and specifically violated by two adjacent MLHs as in Tableau 4.1.[43] This markedness constraint outranks MAX-LINK (3b), which represents a group of faithfulness constraints prohibiting any delinking of tonal feature autosegments in the output (e.g. Yip 2002) to allow the mapping between the input /MLH-MLH/ to the target output [MH-MLH].

(3) a. OCP-MLH: Two adjacent MLH contour tone units are prohibited.
b. MAX-LINK: An input autosegment must have an output correspondent.

/MLH-MLH/	OCP-MLH	MAX-LINK
☞MH-MLH		*
MLH-MLH	*!	

Tableau 4.1. Tone 3 Sandhi triggered by top-ranked OCP-MLH

In line with Zhang (2002) and Zhang & Lai (2010), HS is assumed to be triggered by considerations of ease of articulation in a phonetically shorter context (here non-phrase-final positions). The markedness constraints *NONFINAL-MLH (4) is thus proposed to forbid the disfavored mismatch between the phonetically shorter context and the phonetically longer tone as in Tableau 4.2.

(4) *NONFINAL-MLH: MLH in a non-phrase-final position is prohibited in the output.

/MLH-T/	*NONFINAL-MLH	MAX-LINK
☞ML-T		*
MLH-T	*!	

Tableau 4.2. Half Sandhi triggered by top-ranked *NONFINAL-MLH; T = any tone

Since only non-final MLHs are prohibited, the markedness constraint *MLH, which forbids MLHs in general, must be lower-ranked than the faithfulness

[43] Hyman (1975) propose a synchronic account for T3S based on the phonetic tendency of minimizing tonal ups and downs. Since the full analysis of T3S is not the goal aimed in this chapter, the oversimplified OCP analysis is adopted. This approach to T3S has been criticized for the lack of explanations for why OCP targets only at adjacent MLHs (e.g. Duanmu 1994), and the analysis here does not seek to settle the debate by demonstrating why OCP-MLH outranks other OCP constraints or why other OCP constraints may simply be absent.

constraint to MAX-LINK preserve a phrase-final MLH as in Tableau 4.3.

(5) *MLH: MLH is prohibited in the output.

/T-MLH/	MAX-LINK	*MLH
☞T-MLH		*
T-ML	*!	

Tableau 4.3. Phrase-final MLH is preserved with lower-ranked *MLH; T = any tone

The top-ranked *NONFINAL-MLH also guarantees that the input /MLH-MLH/ does not surface as [MLH-MH] after T3S since this output is harmonically-bounded to the target output [MH-MLH] by violating the same constraints plus *NONFINAL-MLH (Tableau 4.4). Another possible T3S output [MH-MH] is ruled out as the phrase-final MLH is preserved with the higher-ranked MAX-LINK.

/MLH-MLH/	OCP-MLH	*NONFINAL-MLH	MAX-LINK	*MLH
☞MH-MLH			*	*
MH-MH			**!	
MLH-MH		*!	*	*
MLH-MLH	*!			

Tableau 4.4. Other Tone 3 Sandhi outputs ruled out by higher-ranked *NONFINAL-MLH and MAX-LINK

There are other choices for the output of T3S, including changing the first Tone 3 into its half-tone variant (i.e. /MLH-MLH/→[ML-MLH]). I propose that this output is prohibited due to a disfavored long tonal lapse sequence (i.e. MLML), which violates the constraint *LONGLAPSE defined in (6). This constraint is an extension of *LAPSE in Zoll (2002), which forbids a sequence of adjacent non-high pitches. This constraint is necessarily top-ranked for the target output MH-MLH to be the winner as shown in Tableau 4.5.

(6) *LONGLAPSE: MLML (or four adjacent [-high] autosegments) sequence is prohibited in the output.

/MLH-MLH/	OCP-MLH	*LONGLAPSE	MAX-LINK
☞MH-MLH			*
ML-MLH		*!	
MLH-MLH	*!		

Tableau 4.5. Output ML-MLH prohibited with a top-ranked *LONGLAPSE

The introduction of *LONGLAPSE rules out only one possible output of T3S, since the first Tone 3 can also surface as [H] or [HL] to avoid violating OCP-MLH. A perceptual account following Steriade's (2001b) P-Map theory explains why these options are less preferred: [H] and [HL] are perceptually more distant from /MLH/ than [MH]. Huang (2001), Liu & Samuel (2004), and many others have shown that Tone 2 (MH) and Tone 3 (MLH) are perceptually similar and [MH] is thus a preferred output of /MLH/. To capture the difference in the input-output perceptual distortion, MAX-LINK is decomposed into MAX-LINK(M) and MAX-LINK(L/M_H) in (7). The former is violated when an input M autosegment is delinked (e.g. MLH→H or MLH→HL), and the latter is violated only when an input L autosegment is delinked between an input M and H autosegments (i.e. /MLH/→[MH]). The proximate perceptual distance between MH and MLH is modeled by the intrinsic ranking MAX-LINK(M) » MAX-LINK(L/M_H). That is, the removal of an input M autosegment results in an output perceptually more distant from MLH. The selection of the optimal output [MH-MLH] over [H-MLH] or [HL-MLH] is illustrated in Tableau 4.6.

/MLH-MLH/	OCP-MLH	MAX-LINK(M)	MAX-LINK(L/M_H)
☞MH-MLH			*
H-MLH		*!	
HL-MLH		*!	
MLH-MLH	*!		

Tableau 4.6. Selecting the optimal output with a shorter perceptual distance from the input

(7) a. MAX-LINK(M): An input M autosegment must have an output correspondent (i.e. *MLH→H and *MLH→HL)
b. MAX-LINK(L/M_H): An input L autosegment between autosegments M and H must have an output correspondent (i.e. *MLH→MH)

Finally, although MH is perceptually similar to MLH, it should be noted that the output of HS is always [ML] rather than [MH]. This output selection of HS suggests that ML is even more perceptually similar to MLH. The perceptual similarity between ML and MLH has been found in Fon et al. (2004), who demonstrated that the initial fall is an important perceptual cue for Tone 3, whereas the rising pitch contour is perceived as the most salient feature of Tone 2. Results of perceptual experiments in Chen & Tucker (2013) also indicate that coarticulated dipping pitches on sonorant onsets are more frequently identified as the onset pitch of Tone 3. The intrinsically lower-ranked MAX-LINK(H/ML_) in (8) can thus be proposed for this IO mapping between /MLH/ and [ML] as in Tableau 4.7. Despite this closer perceptual distance between ML and MLH, Tone 3 never surfaces as [ML] in the T3S context since the top-ranked is *LONGLAPSE violated as seen above. To sum up here, the full constraint ranking for both T3S and HS can be presented as Figure 4.3.

/MLH-T/	*NONFINAL-MLH	MAX-LINK(M)	MAX-LINK (L/M_H)	MAX-LINK (H/ML_)
☞ML-T				*
MH-T			*!	
H-T		*!		
HL-T		*!		
MLH-T	*!			

Tableau 4.7. Tone 2 MH prohibited by MAX-LINK(L/M_H) as the output of Half Sandhi; T = any tone

(8) MAX-LINK(H/ML_): An input H autosegment preceded by autosegments ML must have an output correspondent (i.e. *MLH→ML).

{OCP-MLH, *NonFinal-MLH, *LongLapse}

|

Max-Link(M)

|

Max-Link(L/M_H)

|

Max-Link(H/ML_)

|

*MLH

Figure 4.3. Constraint ranking for Tone 3 Sandhi and Half Sandhi

2. Unexceptional Tone 3 reduction in Taiwan Mandarin

Standard Mandarin is also the official language or used as a primary dialect in many countries other than Mainland China such as Taiwan, Singapore, etc. Although mutually intelligible, Standard Mandarin and its variants in these areas may slightly differ in vocabulary, segment, and tone inventory. This chapter concentrates on an intriguing tonal difference between Standard Mandarin and Taiwan Mandarin. A possible chain-shift tonal change in Taiwan Mandarin is summarized in §2.1, and a PSI-based account for the change from a full Tone 3 to a half-tone in Taiwan Mandarin is proposed in §2.2. Note that Taiwan Mandarin should not be confused with Taiwanese Mandarin, which has been commonly used as an alternative name of the Southern Min (Xiamen) dialect.

2.1 Chain-shift tonal change in Taiwan Mandarin

Taiwan Mandarin is phonologically similar to Standard Mandarin with a tonal inventory of four lexical tones. Among the four lexical tones, Tone 1 and Tone 4 remain unchanged and are realized as a high level tone and a falling tone respectively, but the phonetic investigation in Fon et al. (2004) reveals variation in Tone 2 and Tone 3. As summarized in §1, Tone 2 in Standard Mandarin has an small initial pitch fall which quickly turns into a rising pitch contour after the tonal onset. In Taiwan Mandarin, although the pitch frequency in Tone 2 remains higher than in Tone 3, the turning point from the initial fall to a rising pitch in Tone 2 has been delayed to the midpoint of the entire tonal contour – a concave contour shape more similar to Tone 3 in Standard Mandarin. Tone 3 has two surface variants [ML] and [MLH] in Standard Mandarin as the consequence of HS. However, in Taiwan Mandarin, Fon et al. found that Tone 3 was unexceptionally shortened regardless of prosodic positions (cf. Shih 1988). The sample pitch trace of the two tones in Taiwan Mandarin is provided in Figure 4.4.

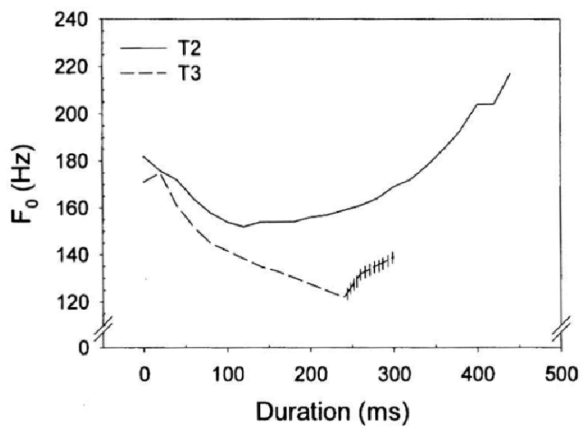


Figure 4.4. Production of [pa] with Tone 2 (T2) and Tone 3 (T3) by a female speaker of Taiwan Mandarin in a carrier sentence *zhege zi shi nian* 'This character is read as ____'; vertical marks indicate creaky voice quality (From Fon et al. 2004:256, Figure 2)

In sum, it seems that Tone 2 and Tone 3 have undergone a chain shift as illustrated in Figure 4.5: Tone 3 has completely transformed into a half-tone, and Tone 2 has adopted the contour shape of the original Tone 3 but still preserves its overall higher pitch frequency (thus is not identical to MLH). The crucial grammar difference between Taiwan and Beijing lies behind the former change of Tone 3, which results in the overall absence of surface [MLH]; *MLH must be higher-ranked than MAX-LINK constraints as in to derive underlying MLHs as MLs in different prosodic position as in Tableau 4.8: The input /MLH-T/ surfaces as [ML-T] as in Beijing Mandarin, but the input /T-MLH/ now always surfaces as [T-ML] with the reduction process in the phrase-final position as well.

Tone 2 (MH) ⟶ Tone 3-like shape
Tone 3 (MLH)
↓
half-tone (ML)

Figure 4.5. A likely chain shift of Tone 2 and Tone 3 in Taiwan Mandarin

/MLH-T/	*NONFINAL-MLH	*MLH	MAX-LINK
☞ML-T			*
MLH-T	*!	*	
/T-MLH/			
☞T-ML			*
T-MLH		*!	

Tableau 4.8. Reduction of /MLH/ in non-final and phrase-final positions with top-ranked *MLH; T = any tone

The assumed constraint ranking for Taiwan Mandarin is illustrated in Figure 4.6, assuming that all markedness constraints are top-ranked but *MLH is intrinsically outranked by *NONFINAL-MLH to demonstrate the inherently disfavored non-final MLHs.

{OCP-MLH, *NONFINAL-MLH, *LONGLAPSE, <u>*MLH</u>}
|
MAX-LINK(M)
|
MAX-LINK(L/M_H)
|
MAX-LINK(H/ML_)

Figure 4.6. Constraint ranking after tonal changes in Taiwan Mandarin; the underlined constraint stands for the crucial ranking difference

2.2 A PSI-based account

The change from a full Tone 3 to a half-tone in Taiwan Mandarin is the focus of the rest of the discussion in this chapter with the hypothesis that the input selection process drove this diachronic development. Under the framework of PSI, three surface variants of Tone 3 [MH] (T3S output), [ML] (HS output), and [MLH] are stored in the lexicon as possible input choices and will be assigned a selection probability (SP) for the probabilistic input selection process. Assume that the constraint ranking in early Taiwan Mandarin is identical to that in Standard Mandarin (see Figure 4.3 in §1). In non-phrase-final position, /ML/ and /MLH/ are neutralized to [ML] due to HS, as shown in Tableau 4.9. That is to say, in this prosodic position, output errors are not generated from both inputs /ML/ and /MLH/ to lower their SP. The MH allomorph can never surface as the target output of HS [ML] since no top-ranked natural constraints forces the input /MH/ to change to [ML] in this context; the target output [ML-T] is thus harmonically bounded by the faithful output [MH-T].

/MLH-T/	*NONFINAL-MLH	MAX-LINK(M)	MAX-LINK (L/M_H)	MAX-LINK (H/ML_)
☞✓ML-T				*
MH-T			*!	
H-T		*!		
HL-T		*!		
MLH-T	*!			
/ML-T/	*NONFINAL-MLH	MAX-LINK(M)	MAX-LINK (L/M_H)	MAX-LINK (H/ML_)
☞✓ML-T				
MH-T			*!	
H-T		*!		
HL-T		*!		
MLH-T	*!			
/MH-T/	MAX-LINK	MAX-LINK(M)	MAX-LINK (L/M_H)	MAX-LINK (H/ML_)
✓MH-T				
☞☹ML-T	*!			

Tableau 4.9. Neutralization between the inputs /ML/ and /MLH/ in the Half Sandhi context; T = any tone, '✓' = correct output, '☹' = output error

The T3S context (i.e. before another MLH) is the only context where all the three allomorphs are neutralized as [MH] since the input /MLH-MLH/ surfaces as [MH-MLH] to satisfy OCP-MLH (Tableau 4.10), /ML-MLH/ surfaces as [MH-MLH] to satisfy *LONGLAPSE (Tableau 4.11), and /MH-MLH/ surfaces faithfully as [MH-MLH] (Tableau 4.12).

/MLH-MLH/	OCP-MLH	Max-Link(M)	Max-Link (L/M_H)	Max-Link (H/ML_)
☞MH-MLH				*
ML-MLH			*!	
H-MLH		*!		
HL-MLH		*!		
MLH-MLH	*!			

Tableau 4.10. From the input MLH to the target output MH in the Tone 3 Sandhi context

/ML-MLH/	*LongLapse	Max-Link(M)	Max-Link (L/M_H)	Max-Link (H/ML_)
☞MH-MLH[44]				
H-MLH		*!		
HL-MLH		*!		
ML-MLH	*!			

Tableau 4.11. From the input ML to the target output MH in the Tone 3 Sandhi context

/<u>MH</u>-MLH/	Max-Link(H/ML_)	*MLH
☞<u>MH</u>-MLH		*
<u>MH</u>-ML	*!	

Tableau 4.12. Input MH surfaces faithfully in the Tone 3 Sandhi context

In a phrase-final position, however, only the input /MLH/ can surface faithfully since there is neither a top-ranked constraint forcing the input /MH/ or /ML/ to change to [MLH] in a phrase-final position. As shown in Tableau 4.13, when Max-Link outranks *MLH, the phrase-final input /MLH/ surfaces faithfully. However, with an underlying phrase-final /ML/ or /MH/, the output [MLH] is harmonically-bounded by the faithful output [ML] or [MH].

[44] The change from the input ML to the output MH violates both lower-ranked Max-Link(L) and Dep-Link(H), which are excluded here to simplify the analysis.

/MLH/	MAX-LINK(H/ML_)	*MLH
☞✓MLH		*
ML	*!	
/ML/		
✓MLH		*!
☞☹ML		
/MH/		
✓MLH		*!
☞☹MH		

Tableau 4.13. Target outputs harmonically bounded by faithful outputs with the inputs /ML/ and /MH/; '✓' = correct output, '☹' = output error

In all three contexts, only the inputs /MH/ and /ML/ may surface as output errors which raise the Individual Error Proportion (IEP) of /MH/ and /ML/ and thus lower SP_{MH} and SP_{ML}. In accord with the standard analysis, the MLH allomorph should be assigned a dominant SP, and Tone 3 remains to be realized as three different allomorphs in different prosodic contexts. However, this prediction disregards the influence from the token frequency of each allomorph of Tone 3. As the phrase length varies, there is always only one single final position but more non-final positions. The ML allomorph, which occurs in a non-final position, must thus have a higher token frequency than the MLH allomorph. The token frequency of the MH allomorph is also lower since the MH allomorph only surfaces when the input has two adjacent Tone 3s.[45] In sum, the ML allomorph should always have the highest token frequency as shown in (9).

(9) General frequency hierarchy of Tone 3 allomorphs
freq(ML) > {freq(MH), freq(MLH)}

The frequency hierarchy may slightly vary across different Tone 3 words because some Tone 3 words may appear more frequently in phrase-final position and some may almost always occur in non-final positions (see also §3.1 below). Assume two Tone 3 words in (10): Although they both are realized as [MLH] in

[45] Based on a syllable-frequency corpus (Da 2004), the type frequency of Tone 3 (254 out of 1,054 syllable types) is only higher than that of Tone 2, and the token frequency of Tone 3 is lowest (14.8% of 192,647,157 syllables). The chance for two adjacent Tone 3s is thus lower than the chance any other di-tonal combinations.

phrase-final positions, the hypothetical word T3A appears mostly in the HS context (i.e. non-final positions) and the hypothetical word T3B is produced in non-final positions less often. The token frequency of the ML allomorph of T3A is thus much higher than that of T3B.

(10) Token frequency in the adult language (Generation 0)
T3A: $freq_{ML}$ >>>>>>>>>>> $freq_{MH}$ > $freq_{MLH}$[46]
T3A production in phrase-final positions: [MLH]
T3B: $freq_{ML}$ > {$freq_{MH}$, $freq_{MLH}$}
T3B production in phrase-final positions: [MLH]

The diachronic tonal change then starts with the two different frequency distributions in the learning inputs. Learners of the next generation (Generation 1) will perceive much more ML allomorph tokens for T3A, and this extreme frequency bias may not be fully countered by the number of output errors generated from the allomorph in a phrase-final position (i.e. [ML], see Tableau 4.13); SP_{ML} is now as high as SP_{MLH} for T3A as shown in (11). By contrast, since the ML allomorph of T3B does not have a significantly higher frequency than the other two allomorphs, the MLH allomorph is still assigned as a dominant SP by producing more targets and fewer errors. [ML] and [MLH] now are in free variation for T3A in phrase-final positions (by selecting either /ML/ or /MLH/ as the input), but [MLH] is still consistently produced for T3B in the same context (by selecting only /MLH/ as the input). Since the outputs [ML] and [MLH] are faithfully derived from the inputs /ML/ and /MLH/, the learners still acquire the target grammar in which MAX-LINK outranks *MLH.

[46] Since T3S also changes a non-final Tone 3 to the MH allomorph, Tone 3 words that occur more frequently in non-final positions, like T3A, should undergo T3S more often, and $freq_{ML}$ is thus also higher than $freq_{MLH}$ for these Tone 3 words (see also the distributional frequencies in Table 4.1 in §3.1).

(11) SP and grammar acquired by the learners of Generation 1

T3A: $SP_{ML} \approx SP_{MLH}$ >>>>>>>>>> SP_{MH}

T3A production in phrase-final positions: [ML], [MLH]

T3B: SP_{MLH} >>>>>>>>>>> $SP_{MH} \approx SP_{ML}$

T3B production in phrase-final positions: [MLH]

Acquired grammar: MAX-LINK » *MLH

When the productions of the above learners become the learning inputs for the next generation (Generation 2), the learning targets are different for T3A and T3B; in phrase-final positions, T3A varies between [ML] and [MLH], and T3B is still [MLH]. When producing T3A in phrase-final positions, if /MLH/ is selected as the input but the target output is [ML], the learners will have to rank *MLH higher than MAX-LINK to produce the target output as in Tableau 4.14. This ranking nevertheless contradicts the grammar required to faithfully derive the target [MLH] from the input /MLH/ of T3B and forces the outputs of T3B to be the output error [ML] as well. With the output error, IEP(MLH) of T3B raises to lower SP_{MLH} of T3B, and SP_{ML} rises for its intrinsically higher token frequency. The changes in the SPs and grammar are summarized in (12): *MLH and MAX-LINK move close to each other due to the conflicting rankings, and SP_{MLH} and SP_{ML} of T3B also become similar.

T3A /MLH/ Target: ML	*MLH	MAX-LINK
☞✓ML		*
MLH	*!	
T3B /MLH/ Target: MLH	*MLH	MAX-LINK
☞☹ML		*
✓MLH	*!	

Tableau 4.14. /MLH/→[ML] with the ranking *MLH » MAX-LINK

(12) SP and grammar acquired by the learners of Generation 2

T3B: $SP_{MLH} \approx SP_{ML}$ >>>>>>>>>>>> SP_{MH}

T3B production in phrase-final positions: MLH, ML

Acquired grammar: MAX-LINK ≈ *MLH

For learners of the following generations, the target output of T3B, like the target of T3A, also varies between [ML] and [MLH] in phrase-final positions for two reasons: First, a higher SP(ML) of T3A allows /ML/ to be selected as the phrase-final input of T3A and surface faithfully as [ML]. Second, because *MLH and MAX-LINK are moved closer, *MLH may outrank MAX-LINK at times to derive the input /MLH/ as the output [ML]. With an even higher number of /MLH/→[ML] mappings, the pressure for these new learners to rank *MLH higher than MAX-LINK becomes stronger. Eventually, *MLH will fully dominate MAX-LINK, and both Tone 3 words are consistently produced as [ML]. At this point, the shift from a full concave tone to a half-tone is complete.

To sum up here, the PSI account assumes the following steps to complete the diachronic development. First, some Tone 3 words occur in non-final positions more frequently and surface mostly as [ML]. Such a high token frequency then allows /ML/ to be selected as the input of these Tone 3 words and surface faithfully as [ML] even in phrase-final positions. For these Tone 3 words, [ML] is at times recognized as the target tone regardless of prosodic positions, which requires *MLH to outrank MAX-LINK constraints. As more /MLH/→[ML] demands the ranking *MLH » MAX-LINK, *MLH will eventually fully dominates MAX-LINK, and other Tone 3 words are consistently produced as [ML] as well.

Assuming that the diachronic development is in progress in Standard Mandarin and the grammar change has not completed, the distributional frequency in different prosodic positions largely determines the change pace from [MLH] to [ML]. In the above example, I have demonstrated that the Tone 3 word that occurs in phrase-final positions more frequently may change its phonetic realization to ML slower due to a higher token frequency of its MLH allomorph. It is thus predicted that with a group of different Tone 3 words, a linear correlation should be found between their phrase-final token frequency and the chance for their [MLH] realization to be preserved on the surface. In the rest of this chapter, this hypothesis will be tested with an elicitation experiment in §3, which is later modeled by PSI-OT-GLA in §4.

3. Testing Tone 3 production in Standard Mandarin

The above sections have spelled out the hypothesis of how Tone 3 was transformed from a full concave tone to a half-tone in Taiwan Mandarin by selecting a high-frequency half-tone allomorph as the input of Tone 3 words. This hypothesis nevertheless cannot be tested in Taiwan Mandarin since this sound change seems to have already run its course in this dialect. If the hypothesis is true, however, Standard Mandarin should be undergoing the same diachronic development as half-tone allomorphs also tend to have a higher token frequency and are thus more likely to be selected as Tone 3 word inputs. An elicitation experiment was thus designed to examine the production of Tone 3 words in the phrase-final context by native Standard Mandarin speakers.

3.1 Materials

The elicitation task includes eighteen Tone 3 words as listed in Table 4.1. These Tone 3 words were also target words in the elicitation task in Fon et al. (2004), and were produced with a half-tone in phrase-final positions in Taiwan Mandarin. The same words are thus selected for the current elicitation task to replicate the same result in Standard Mandarin.

The token frequency of each word in different contexts (i.e. non-final position, phrase-final position, T3S context) is calculated from The Lancaster Corpus of Mandarin Chinese (McEnery & Xiao 2003-2008). The contextual distributional probabilities are calculated to illustrate the chance for a Tone 3 word to be produced in non-final and phrase-final position. The Ratio of Phrase-final Use (RPU) of each Tone 3 word appearing in phrase-final positions is also calculated following the formula in (13) as an indicator of the phrase-final preference: A higher RPU stands for a stronger tendency for a Tone 3 word to appear in phrase-final positions.

(13) Ratio of Phrase-final Use is defined as the following formula, where freq(final) stands for the token frequency of a Tone 3 word in phrase-final positions and freq(total) stands for the total token frequency of the Tone 3 word.

$$RPU = \frac{freq(final)}{freq(total)}$$

Tone 3 Word	Non-Final	Phrase-Final	T3S Context	Total	RPU
礼 [li^{213}] ‘gift’	19 (0.6%)	25 (8.3%)	0	44	0.568
友 [iou^{213}] ‘friend’	9 (0.3%)	8 (2.6%)	0	17	0.471
眼 [iẽn^{213}] ‘eye’	82 (2.6%)	54 (17.9%)	4	140	0.386
远 [yãn^{213}] ‘far’	95 (3%)	49 (16.2%)	1	145	0.338
止 [ʐɨ^{213}] ‘stop’	12 (0.4%)	6 (2%)	0	18	0.333
险 [ɕiãn^{213}] ‘danger’	8 (0.3%)	4 (1.3%)	0	12	0.333
组 [tsu^{213}] ‘group’	52 (1.6%)	16 (5.3%)	0	68	0.235
五 [u^{213}] ‘five’	245 (7.7%)	76 (25.2%)	36	357	0.213
体 [ti^{h213}] ‘body’	80 (2.5%)	23 (7.6%)	6	109	0.211
桶 [t^{h}uɔ̃ŋ^{213}] ‘bucket’	10 (0.3%)	3 (1%)	3	16	0.188
补 [pu^{213}] ‘mend’	15 (0.5%)	4 (1.3%)	3	22	0.182
喜 [ɕi^{213}] ‘happy’	34 (1.1%)	8 (2.6%)	3	45	0.178
请 [tɕ^{h}ĩŋ^{213}] ‘invite’	231 (7.3%)	8 (2.6%)	53	292	0.027
俩 [lia^{213}] ‘pair’	71 (2.2%)	2 (0.7%)	1	74	0.027
产 [tʂ^{h}ãn^{213}] ‘yield’	41 (1.3%)	1 (0.3%)	1	43	0.023
指 [ʐɨ^{213}] ‘point’	319 (10.1%)	5 (1.7%)	8	332	0.015
给 [kei^{213}] ‘give’	803 (25.4%)	6 (2%)	304	1113	0.005
使 [ʂɨ^{213}] ‘make’	1038 (32.8%)	4 (1.3%)	237	1279	0.003
Total	3164	302	660	4124	

Table 4.1. Contextual token frequency of Tone 3 words selected for the elicitation experiment; parenthesized proportions represent the distribution probability of a Tone 3 word in different contexts; non-final = non-final token frequency, Final = phrase-final token frequency, T3S Context = T3S token frequency, Total = total token frequency, RPU = Ratio of Phrase-final Use

Note that the corpus is a text corpus rather than a spontaneous speech corpus since the small size of currently available spontaneous speech corpora of Standard Mandarin have a common sampling issue; the distributional frequency of some Tone 3 words cannot be evaluated as these words (lower frequency words in particular) are absent from such speech corpora. With a text corpus, a non-final or phrase-final position can only be defined with the presence of a comma or period. If a Tone 3 word is immediately followed by a comma or period, its prosodic position is treated as phrase-final and the MLH allomorph is assumed to surface. If there is no comma or period following a Tone 3 word, the position is presumably non-final, and the ML or MH allomorph should surface. The risk of this text-based definition on prosodic positions is the overestimation of the number of phrase-final positions;

in real speech, the presence of punctuation is not necessarily realized with a pause, and there could also be more phrase-final particles in real speech than in written texts.

Despite the possibility of overestimation, I suggest that the above frequency count still generally reflects the distributional tendency of different words. For example, Tone 3 words which appear more frequently in a phrase-final position are mostly nouns. The words occurring less frequently in a phrase-final position are mostly transitive verbs and modifiers, which must be followed by other words and thus not phrase-final. The eighteen Tone 3 words, based on the RPUs, can be further categorized as three phrase-final proportion groups as in (14). If the hypothesis is true, we will discover an inverse correlation between RPU and the frequency of a Tone 3 word produced as the ML allomorph in phrase-final positions, e.g. the members of (14c) will be realized as the half-tone ML more frequently than the members of (14a) and (14b) in phrase-final positions. The elicitation task also includes 42 fillers of Tone 1, Tone 2, and Tone 4 words (i.e. fourteen words from each tone type; see Appendix C) to conceal the goal of the task, which join the eighteen Tone 3 words to make a wordlist of $(14 \times 3) + 18 = 60$ words.

(14) Tone 3 words grouped by RPUs

a. High RPU family (0.568~0.333):
礼 (gift), 友 (friend), 眼 (eye), 远 (far), 止 (stop), 险 (danger)

b. Medium RPU family (0.235~0.178):
组 (group), 五 (five), 体 (body), 桶 (bucket), 补 (mend), 喜 (happy)

c. Low RPU family (0.027~0.003):
请 (invite), 俩 (pair), 产 (yield), 指 (point), 给 (give), 使 (make)

3.2 Procedure

The elicitation tasks were conducted inside a sound-attenuated booth in the Alberta Phonetics Laboratory at the University of Alberta. There were three elicitation tasks, which collected the production of Tone 3 words in non-final and phrase-final positions respectively. In the first elicitation task, participants were asked to put every word in a phrase-final position in the carrier sentence '[tɕiãŋ^{24} ʂãŋ^{51} kʰɤ^{55} tʂɤ^{0} ___] (Inscribed on the wall is the word ___)'. In the second task, they were asked to produce each word in a phrase-initial position in the carrier sentence '[___ kʰɤ^{55} tsai^{51} tɕiãŋ^{24} ʂãŋ^{51}] (the word ___ is inscribed on the wall)'. Finally, each word was produced in isolation. The goal of the first task was to observe the frequency of the ML or MLH allomorph of each Tone 3 word in a phrase-final position. The goal of the second task was to confirm that Tone 3 was realized as a half-tone [ML] without exception in a non-final position. The last task was designed to verify the production of Tone 3 words in isolation; if Tone 3 words are always produced as the ML allomorph in isolation by a speaker, the diachronic change may be complete for this speaker.

Before the recording sessions, participants were told that they would have to read each simplified Chinese character appearing in the middle of a computer screen with or without the carrier sentences above. They were specifically told that the experiment was not designed to evaluate the precision of their Standard Mandarin pronunciation, and that they could speak normally as in daily conversation. Before each of the first two recording sessions, the participants were then asked to listen to four example phrases produced by a male native Mandarin speaker. The example phrases were recorded at a normal speech rate with a mean length of 2.17 s (*sd* = 0.35 s). The phrase-initial and phrase-final target words have a mean length of 381.8 ms (*sd* = 38.8 ms) and 256.8 ms (*sd* = 44.9 ms). The example phrases were also produced without any pause between syllables. The above instruction was designed to prime the participant with normal speech tokens, which might have helped avoid hyperarticulation in careful speech in the recording sessions. Note that due to this possible priming effect, Tone 3 words were not included in the example phrases to prevent the production of Tone 3 words from being influenced by the example tokens.

(15) a. Four example phrases for Session I:

i. tɕiãŋ^{24} ʂãŋ^{51} kʰɤ^{55} tʂɤ^{0} <u>tʂau^{51}</u>
"inscribed on the wall is the word 'illuminate.'"

ii. tɕiãŋ^{24} ʂãŋ^{51} kʰɤ^{55} tʂɤ^{0} <u>pʰei^{24}</u>
"inscribed on the wall is the word 'accompany.'"

iii. tɕiãŋ^{24} ʂãŋ^{51} kʰɤ^{55} tʂɤ^{0} <u>fɔ̃ŋ^{55}</u>
"inscribed on the wall is the word 'fruitful.'"

iv. tɕiãŋ^{24} ʂãŋ^{51} kʰɤ^{55} tʂɤ^{0} <u>tuei^{51}</u>
"inscribed on the wall is the word 'correct.'"

b. Four example phrases for Session II:

i. tʂau^{51} kʰɤ^{55} tsai^{51} tɕiãŋ^{24} ʂãŋ^{51}
"the word 'illuminate' is inscribed on the wall."

ii. pʰei^{24} kʰɤ^{55} tsai^{51} tɕiãŋ^{24} ʂãŋ^{51}
"the word 'accompany' is inscribed on the wall."

iii. fɔ̃ŋ^{55} kʰɤ^{55} tsai^{51} tɕiãŋ^{24} ʂãŋ^{51}
"the word 'fruitful' is inscribed on the wall."

iv. tuei^{51} kʰɤ^{55} tsai^{51} tɕiãŋ^{24} ʂãŋ^{51}
"the word 'correct' is inscribed on the wall."

In each trial during the formal recording sessions, a word was randomly selected from the 60-word list and presented in the middle of a computer screen. Each word was randomly chosen three times, resulting in a total of 60 × 3 = 180 trials for each recording session. Participants had three seconds in each trial to produce the phrase, and the experiment automatically moved to the next trial after the time limit. If the participants could not recognize the character in a trial, they responded with '[pu^{24} huei^{51}] (I don't know)', and the recording was excluded from the results. There was a five-second break for every ten trials for storing audio files on the experiment running computer.

The target phrases were recorded using a CountryMan® EP6 head-mounted microphone via an Alesis MultiMix® 8 USB FX mixer directly output to Korg® MS 2000S recorder at a sampling frequency of 44,100 Hz with a 16-bit sampling rate.

3.3 Participants

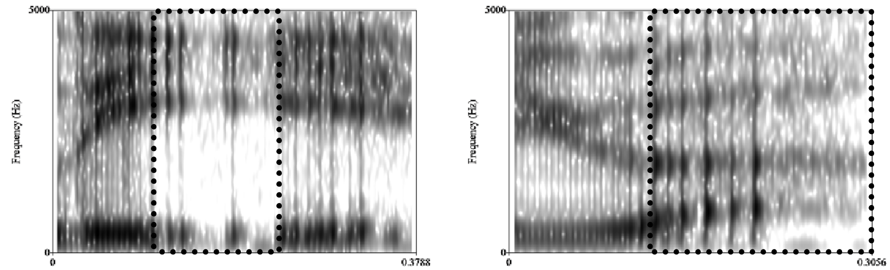
Fourty participants were recruited at the University of Alberta, who were enrolled as undergraduate students at the time of the study. One was born in Edmonton, Alberta, Canada. Although Standard Mandarin is his first language, his reading training was not sufficient to allow him to read Chinese characters fluently through the experiment. All remaining participants were born in China as native speakers of Standard Mandarin. After excluding the single exception, only results from 39 participants are included in the following analysis and discussion. Some of the 39 participants could speak other Chinese dialects, and the dialect effect will be evaluated in the discussion of the results below. None of the participants reported any language impairment issue.

3.4 Results

The result of three participants were excluded from the following discussion for the following reasons: First, two participants produced Tone 2 and Tone 3 very similarly presumably as a result of an ongoing tonal merger process caused by the perceptual similarity between Tone 2 and Tone 3.[47] Second, in the first task, one participant produced sentences with significantly different speech rates (an average length of 1,166 ms for the first ten sentences (*sd* = 67 ms) and 931 ms for the last ten sentences (*sd* = 45 ms)). Therefore, only results from remaining 36 participants are covered in the upcoming discussion. Aside from these three individual exceptions, tokens with accidents like incorrect pronunciations, laughs, coughs, hesitations, yawns, silences, etc., were also excluded from the results (reported below).

The judgment on the production of Tone 3 words mainly relied on visual inspection of spectrograms in Praat (Boersma & Weenick 2013); the participants' voice might have been unexpectedly weak, which made impressionistic transcription difficult to decide whether Tone 3 words were produced as the ML or MLH allomorph. F0 trace is most indicative in judging the realization of tones, but the F0 tracker in Praat frequently failed to generate a complete F0 contour of Tone 3 words due to a weak voice or irregular glottal pulses. Two alternative visual cues helped determine the contour shape of these Tone 3 words. First, since the valley of Tone 3 is frequently associated with creaky voice (e.g. Hockett 1947, Garding et al. 1986), if a modal-creaky-modal voice sequence is visible as in Figure 4.7a, the Tone 3 word is produced as a full concave contour tone. If as in Figure 4.7b, a modal voice is only followed by a creaky voice, the Tone 3 word is produced as the ML allomorph.

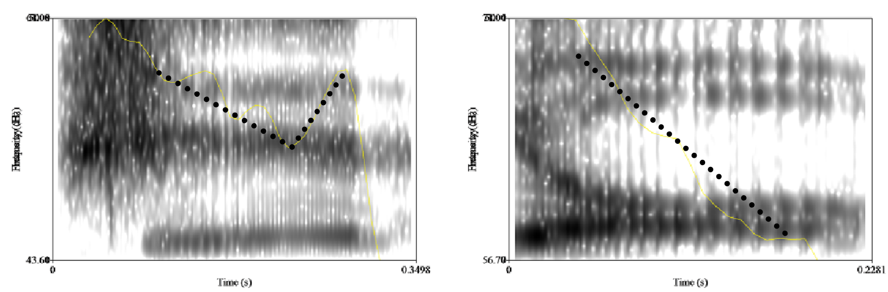
[47] This assumption is a potential explanation for why Tone 2 has a contour shape more similar to Tone 3 in Taiwan Mandarin (see §2.1). This issue is open for further study.



(a) (b)

Figure 4.7. (a) MLH of Tone 3 words represented by modal-creaky-modal sequence and (b) ML of Tone 3 words represented by modal-creaky sequence; boxes with dotted lines isolate areas of creaky voice

The second visible cue for the distinction between a full concave contour and a half-tone is the amplitude contour. Fu & Zeng (2000), Liu & Samuel (2004), Whalen & Xu (1992), and among many others report a direct correlation between the pitch height and the amplitude; a higher pitch is associated with a higher amplitude. Therefore, a full concave pitch contour is realized along with a concave amplitude contour as in Figure 4.8a, and a ML pitch contour coexists with a falling amplitude contour as in Figure 4.8b.



(a) (b)

Figure 4.8. (a) MLH of Tone 3 words represented by a concave amplitude contour and (b) ML of Tone 3 words represented by a falling amplitude contour; yellow lines represent the amplitude contours generated in Praat, and dotted lines represent the general shape of amplitude contours

In the second task, it was confirmed that Tone 3 words are produced as half-tones without exception; the phonological constraint forbidding non-final concave tones is never violated in Standard Mandarin as shown in previous studies. Both of the first and third tasks required the speakers to produce Tone 3 words in phrase-final positions, but in the third task Tone 3 words were produced in isolation and mostly realized as a full concave tone. I arbitrarily set the variation threshold to 5% of the total Tone 3 word tokens (i.e. $18 \times 3 \times 0.05 = 2.7$); that is, if a speaker produced fewer than 3 tokens different from other tokens (e.g. either 2 [ML]s vs. 52 [MLH]s or 2 [MLH]s vs. 52 [ML]s), the Tone 3 word production of this speaker was categorized as lack of variation. As a result, 29 out of the 36 participants did not show variations in the third task; 27 speakers produced almost every Tone 3 word as [MLH] in isolation, and two speakers almost always produced [ML] for Tone 3 words in this context. The reason for the lack of variation in the isolation context might be as follows: When Tone 3 words were produced in isolation, the participants easily emphasized each Tone 3 words and produced the full concave variant, which was still considered as the standard pronunciation of the citation form of Tone 3 words. Interestingly, among the 27 participants, despite constantly producing [MLH] for Tone 3 words in isolation, two participants (Subject 4 and 30) almost always produced the half-tone variants in the first task (only one and two tokens of the full concave variants were produced by each participant). This sharp between-task divergence demonstrates different speech modes activated in different contexts. Since the production in the hopefully more natural speech mode is of current interest, the following discussion will mainly focus on the results of the first task with an experiment setting for spontaneous speech (i.e. producing target words in a phrase-final position with a carrier sentence).

The production of Tone 3 in phrase-final positions in the first task exhibits more surface variations as predicted, although some speakers were still constantly producing almost the same outputs. After applying the same variation threshold, twelve participants were considered not producing surface variations of Tone 3 words. Among the twelve participants, eight of them produced the majority of the full concave variants, presumably emphasizing the Tone 3 words as in the third task. The other four participants produced almost no full concave variants for mixed reasons, and one possible line is that the shift in the Tone 3 production has been complete for these speakers. It is worth noting that the eight speakers who produced variations of Tone 3 words in the third task also produced variations in the first task.

The variations produced by the 24 participants are shown as the word-specific proportion of the full concave variant in Table 4.2. Out of $54 \times 24 = 1{,}296$ tokens, 71 tokens (5.4%) were excluded due to recording accidents mentioned earlier. Subject

1 and Subject 32 both had a highest but seemingly acceptable number of eight excluded tokens (8/54 = 14.8%) and their results were thus still analyzed. Although it is not immediately obvious whether the proportion significantly drops with the decreasing RPU, the Tone 3 words 礼 and 友 were indeed produced as the full concave variant more frequently than 给 and 使.

Tone 3 Word		MLH proportion	RPU	Total Token Frequency
礼 [li^{213}]	'gift'	57.1%	0.568	44
友 [iou^{213}]	'friend'	54.3%	0.471	17
眼 [iẽn^{213}]	'eye'	37.1%	0.386	140
远 [yãn^{213}]	'far'	44.9%	0.337	145
止 [zɨ^{213}]	'stop'	50%	0.333	18
险 [ɕiãn^{213}]	'danger'	46.5%	0.333	12
组 [tsu^{213}]	'group'	38.6%	0.235	68
五 [u^{213}]	'five'	62.9%	0.212	357
体 [tʰi^{213}]	'body'	52.2%	0.211	109
桶 [tʰuɔ̃ŋ^{213}]	'bucket'	45.3%	0.188	16
补 [pu^{213}]	'mend'	53.1%	0.182	20
喜 [ɕi^{213}]	'happy'	53.6%	0.178	45
请 [tɕʰĩŋ^{213}]	'invite'	41.8%	0.027	292
俩 [lia^{213}]	'pair'	45.6%	0.027	74
产 [tʂʰãn^{213}]	'yield'	35.9%	0.023	43
指 [zɨ^{213}]	'point'	50.7%	0.015	332
给 [kei^{213}]	'give'	46.2%	0.005	1113
使 [ʂɨ^{213}]	'make'	47%	0.003	1279

Table 4.2. By-item proportions of the full concave variant in phrase-final positions (MLH proportion) in the second task; proportions are equal to the number of concave tokens divided by 72 (three tokens for each Tone 3 word from 24 participants). RPU = Ratio of Phrase-final Use

3.5 Analysis

The production tokens of Tone 3 words were labeled either as full concaves or as half-tones, and the set of binary data was analyzed using Linear-Mixed Effects Logistic Regression package (lme4; Bates et al. 2013) in R (R Team 2012) with two main predictors RPU (Ratio as the variable name) and logFreq (log-transformed total frequency). The two seemingly unrelated predictors are negatively correlated

(r = -0.54, df = 16, p = 0.021), which may lead to a multicollinearity problem in a regression model. The variation inflation factor (VIF) of each predictor is thus calculated to evaluate the amount of variance added to an estimated regression coefficient by multicollinearity; a high VIF (> 5) suggests a significantly increased variance and thus a severe multicollinearity effect. Subject was included as a random effect, and the random slope of each predictor is also controlled (Barr et al. 2013).

The hypotheses of the Tone 3 word production are listed in (16). If neither predictor is significant, the null hypothesis (16a) is true: The selection between the full concave and the half-tone variants is random. If Ratio is significant, Hypothesis 1 (16b) is true as predicted by the PSI account: The more frequently a Tone 3 word occurs in phrase-final positions, the more likely its full concave variant is selected and produced. Hypothesis 2 (16c) is based on the observation that high frequency words are more inclined to undergo phonetic reduction (e.g. Bybee 2001, Pierrehumbert 2002). If Tone 3 words with a higher total frequency were produced as the reduced variants (i.e. half-tone) more frequently, Hypothesis 2 (16c) is true.

(16) Hypotheses of the production of Tone 3 words
  a. Hypothesis 0: The choice between the full concave and the half-tone variants is purely by chance.
  b. Hypothesis 1: The proportion of the full concave variants is directly related to Ratio.
  c. Hypothesis 2: The proportion of the full concave variants is inversely related to logFreq.

The result of the regression analysis is illustrated in Table 4.3, in which the main effect of Ratio, although non-significant, is trending and thus opens the possibility that Hypothesis 1 is true; the production of the full concave variant can be linked to the phrase-final frequency of Tone 3 words. On the other hand, the effect of logFreq is far above the significance level of 0.05, suggesting a weak connection between the chance of producing a full concave variant and a log token frequency. Low VIFs (< 5) indicate that the current regression model is not severely undermined by the multicollinearity issue. The percentages of the full concave variants are plotted against the ratios with the predicted decreasing tendency toward low-ratio Tone 3 words in Figure 4.9.

Fixed effects:

	Estimate	Std. Error	z-value	Pr(>|z|)	VIF
(Intercept)	-0.49087	0.61055	-0.804	0.4214	
Ratio	0.99630	0.55359	1.8	0.0719	1.536227
logFreq	0.04384	0.06715	0.653	0.5139	1.536227

Table 4.3. Linear Mixed-Effect Regression analysis of the Tone 3 word production

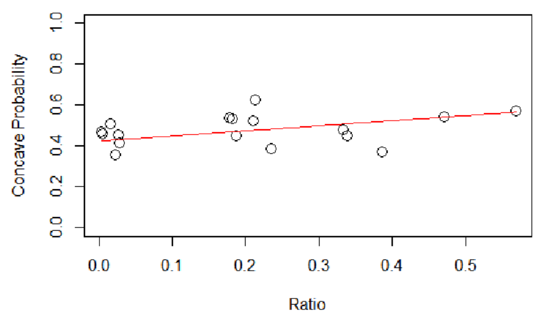


Figure 4.9. Predicted decreasing tendency in the proportion of the full concave variants

The second regression analysis further included the binary predictor Dialects (Yes/No) to examine its interaction with Ratio due to a possible language contact effect which prompted the speakers to produce fewer full concave variants. For example, Cantonese has a low-falling tone similar to the half-tone variant in Standard Mandarin but no full concave tone. If the speakers whose Mandarin production was affected by their knowledge of Cantonese, they might produce more half-tone variants.[48] The result in Table 4.4 does not demonstrate any drastic change the output as the main predictor Ratio remains trending, and the independent variable Dialect and its interaction with Ratio are not significant.[49] The effect of Ratio on the production of Tone 3 words is thus not specific to speakers of other Mandarin dialects.

[48] The influence from Standard Mandarin on minor dialects might be more common as Standard Mandarin is more prestigious as an official language. For example, Zhang & Liu (2011) found a strong influence of Standard Mandarin on the tonal production of the Tianjin dialect.

[49] The two regression models were compared within an ANOVA, which shows that the second model does not significantly improve the predictability from the first model ($X^2 = 3.65$, $df = 6$, $p = 0.72$) with a higher AIC (1268.6 vs. 1276.9) and BIC (1314.5 vs. 1353.5).

Fixed effects:

	Estimate	Std. Error	z-value	Pr(>|z|)	VIF
(Intercept)	-0.90630	0.61009	-1.486	0.1374	
Ratio	1.15559	0.65912	1.753	0.0796	2.092629
DialectYes	1.09869	0.74346	1.478	0.1395	1.207045
logFreq	0.04758	0.06788	0.701	0.4833	1.503091
Ratio:DialectYes	-0.51401	0.97192	-0.529	0.5969	1.822612

Table 4.4. Linear Mixed-Effect Regression analysis of the Tone 3 word production with the additional predictor Dialect

Finally, although there were only seven participants producing similar variations in the third task (i.e. producing words in isolation), their performance in this context can still serve as an indicator of whether the Ratio effect is consistent. The production data was thus included in a post-hoc analysis with the two identical main predictors Ratio and logFreq as shown in Table 4.5. Unlike in the previous analysis, both predictors Ratio and logFreq are highly significant. This result nevertheless does not verify Hypothesis 2 since it demonstrates a direct relation between word frequencies and concave percentages, which contradicts the assumption that high-frequency Tone 3 words undergo half-tone reduction more easily. The more important difference from the previous results is that the effect of Ratio on concave percentages becomes even stronger. While the result here needs to be interpreted with caution since experimental results from only seven participants are less indicative, it seems reasonable to speculate that the trending Ratio effect in the first task and the significant Ratio effect in the third task may not be sheer coincidence.

Fixed effects:

	Estimate	Std. Error	z-value	Pr(>|z|)	VIF
(Intercept)	-2.308	1.0245	-2.253	0.02427 *	
Ratio	3.7325	1.1132	3.353	0.0008 **	1.559767
logFreq	0.4328	0.1452	2.98	0.00289 **	1.559767

* $p \leq 0.05$, ** $p \leq 0.01$

Table 4.5. Linear Mixed-Effect Regression analysis of the Tone 3 word production of the seven speakers showing variations in the third task

3.6 Discussion

The experimental results confirmed the same contour simplification development, which has been completed in Taiwan Mandarin, is under way in Standard Mandarin. More importantly, the analyses also suggest an independent diachronic developmental progress for each Tone 3 word with a different non-final and phrase-final frequency as predicted by the PSI account. The trending positive correlation between Ratio and the concave percentages is nevertheless weaker than expected, particularly because the Tone 3 words with an extremely low ratio were still frequently produced as the full concave variant (see Table 4.2 and Figure 4.9). As mentioned earlier, the design of the first and third tasks might not be able to consistently activate speakers' natural speech mode. In order to control the prosodic context in the first task, the same carrier sentence was used for each trial. Some speakers may have treated the elicitation task as a routine procedure and concentrated only on replacing the phrase-final word without actually accessing lexical information other than the emphatic pronunciation. Without any speech context in the third task, the production only became more single-toned. The current results may be sufficient to shed some light on the ongoing diachronic change in Standard Mandarin, but a design with a more natural speech scenario is desired to allow the emergence of a stronger Ratio effect on the Tone 3 word production. More accurate non-final vs. phrase-final frequency counts from a spontaneous speech corpus should lead to a more solid analysis as well.

Another intriguing question would be, if Standard Mandarin is undergoing exactly the same diachronic development discovered in its Taiwan variant, why does the changing progress in the former seem far behind the developmental stage in the latter? Part of the story might be the effect of language contact. In Taiwan, southern cities in particular, the language population of the Southern Min (or Xiamen) dialect is also large and still frequently acquired as the first language. Complex contour tones (i.e. concave and convex) are absent in the Southern Min dialect, and its low level tone is acoustically akin to the half-tone in Standard Mandarin (see Lin 1988 and Peng 1997). This resemblance may drive speakers to borrow the low-level production in the Southern Min dialect into Standard Mandarin and replace the concave production of Tone 3 words, accelerating the Tone 3 change in Taiwan. This line is open for future study of phonetic investigations into the production of Tone 3 words by older Standard Mandarin monolinguals and younger bilinguals of Standard Mandarin and the Southern Min dialect.

4. Modeling the Tone 3 change with PSI-OT-GLA

Following the above experimental results, PSI-OT-GLA was modified to capture the diachronic development of Tone 3 in Standard Mandarin as a result of recognizing a different basic allomorph. The first change is in the structure of learning inputs as illustrated in §4.1. The second change lies in the constraint grammar as shown in §4.2, which now includes intrinsic constraint rankings. The generation of output candidates from random input phrases will also be discussed in the same section. The change in the implementation and calculation of lexical factors as in previous simulations will be addressed in §4.3. The complete learning process in every learning cycle after the above modifications will be illustrated in §4.4. Finally, to model diachronic changes, PSI-OT-GLA was re-designed to be able to convert the learning results of one generation into the learning inputs for the next generation as shown in §4.5. The following simulation was implemented via Java® programming in Eclipse Standard version 4.3.2 (The Eclipse Foundation 2014) as in previous chapters.[50]

4.1 Input structure

Unlike in the simulation of the Dutch morphophonological acquisition in §3, learning inputs are no longer individual lexical items but random phrases for Standard Mandarin learners, and the number of phrases in a language is infinite. To simulate the generation of random phrases, the following procedures have been adopted. First, the number of words in each learning input is randomly selected between 1 and 5 words. Second, the tone of each word of the input phrases is determined upon the distributional frequency of the four lexical tones calculated from the syllable corpus (Da 2004): 16.7% for Tone 1 words, 18.4% for Tone 2 words, 14.8% for Tone 3 words, and 42.5% for Tone 4 words. If it is a Tone 3 word, the probability of selecting one of the eighteen target Tone 3 words is equal to the token frequency proportion in different prosodic contexts repeated as Table 4.6 below; some Tone 3 words are produced more often in a non-final position, and some appear more frequently in a phrase-final position.

[50] The source code is available at http://hdl.handle.net/10402/era.39160.

	礼[li^{213}]	友[iou^{213}]	眼[iẽn^{213}]	远[yãn^{213}]	止[ʐɨ^{213}]	险[ɕiãn^{213}]
Final	8.3%	2.6%	17.9%	16.2%	2%	1.3%
Non-final	0.6%	0.3%	2.6%	3%	0.4%	0.3%
	组[tsu^{213}]	五[u^{213}]	体[tʰi^{213}]	桶[tʰuõŋ^{213}]	补[pu^{213}]	喜[ɕi^{213}]
Final	5.3%	25.2%	7.6%	1%	1.3%	2.6%
Non-final	1.6%	7.7%	2.5%	0.3%	0.5%	1.1%
	请[tɕʰĩŋ^{213}]	俩[lia^{213}]	产[tʂʰãn^{213}]	指[ʐɨ^{213}]	给[kei^{213}]	使[ʂɨ^{213}]
Final	2.6%	0.7%	0.3%	1.7%	2%	1.3%
Non-final	7.3%	2.2%	1.3%	10.1%	25.4%	32.8%

Table 4.6. Frequency proportions (i.e. distributional probabilities for Tone 3 word selection) of the 18 target Tone 3 words in different contexts

When the input selection process is active in the M stage simulation, the input of the selected Tone 3 word will be chosen from the word's stored allomorphs. The construction of a learning input of a two-word phrase is illustrated in Figure 4.10. The algorithm first determines which lexical tone represents which word, and if it is a Tone 3 word, the algorithm probabilistically selects one of the eighteen Tone 3 words. The last step is to select the input of the chosen Tone 3 word from its three stored allomorphs.

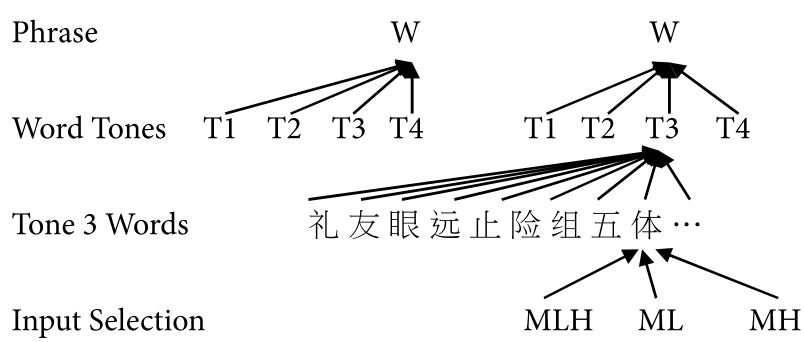


Figure 4.10. Constructing a two-word phrase from selecting the lexical tone of each word to selecting the input of a Tone 3 word

4.2 Constraint grammar, output candidates, and target outputs

The target grammar acquired by the Generation 0 machine speaker is repeated in (17), and a Generation n learner ($n > 0$) is expected to approach this grammar from the initial Markedness » IO-Faithfulness ranking as in (18) via PSI-OT-GLA. Two unnatural markedness constraints in (19) are included in the algorithm by default. The constraint (19a) may be created to derive the target tone ML in a non-final position from the MH allomorph of a Tone 3 word in the input. Likewise, (19b) might emerge in the constraint grammar to produce the phrase-final target output MLH of Tone 3 from the selected MH and ML allomorphs. The above two derivations are not possible with any ranking of natural constraints. As defined in §2.1 of Ch. 2, whether an unnatural markedness constraint is active for its violations to be evaluated during the learning process depends on the number of output types (rather tokens) violating the constraint, based on the Tolerance Principle. Taking (19a) as an example, if there are 100 output types with one or more non-final positions, the number of output types with a non-ML tone in a non-final position must be lower than 100 / log(100) = 21.7 for (19a) to emerge in a constraint grammar.

(17) Target ranking acquired by the Generation 0 speaker
{OCP-MLH, *NONFINAL-MLH, *LONGLAPSE} »
MAX-LINK(M) » MAX-LINK(L/M_H) » MAX-LINK(H/ML_) » *MLH

(18) Initial ranking at the beginning of the learning process
{OCP-MLH, *NONFINAL-MLH, *LONGLAPSE, *MLH} »
MAX-LINK(M) » MAX-LINK(L/M_H) » MAX-LINK(H/ML_)

(19) Two potential unnatural markedness constraints
a. NONFINALDIPPING: In a non-final position, a tone must surface as ML.
b. FINALCONCAVE: In a phrase-final position, a tone must surface as a concave tone.

Note that two intrinsic rankings are involved in the grammar learning process, and one specific issue brought by these rankings is how to preserve these intrinsic rankings in a gradual promotion and demotion process. In the following simulations, the algorithm will force the intrinsically ranked constraints to maintain a difference of at least 5 between each two constraint values in the intrinsic ranking. If, for example, a promotion of an intrinsically lower-ranked constraint lowers the value difference to below 5, the higher-ranked constraint(s) in this intrinsic ranking will be automatically promoted as well to maintain this difference. With this minimal interval, the initial constraint values for the ranking MAX-LINK(M) » MAX-

LINK(L/M_H) » MAX-LINK(H/ML_) is 10 » 5 » 0. Otherwise, markedness constraints have an initial value of 100 and F constraints have an initial value of 0 as in previous simulations. Unnatural markedness constraints, if created, are also initially ranked at the top with an initial value of 100. As in previous simulations, the same constraint demotion bias (see §2.3 of Ch. 2) is integrated into the algorithm, and the promotion plasticity is set to 0.1.

In the modeling of the Dutch morphophonological acquisition, a finite set of output candidates can be generated with fixed constraints and possible inputs, as listed in Appendix B. Due to the random generation process of input phrases, however, it is impractical to enumerate all the possible inputs and their non-harmonically-bounded candidates. To simplify the computational process, some constraints were imposed to help limit the types of output candidates. First of all, the output candidates must include at least the target output and the output fully faithful to the input. Second, when MLH appears in the phrase-final position in the input, candidates which only differ from the target output in the phrase-final tone (surface as ML, MH, and H) will also be generated. Output candidates violating the three top-ranked markedness constraints are redundant since they are never target outputs and winners. Finally, harmonically-bounded output candidates are omitted as usual.

Considering the four-word input /H-MLH-MH-MLH/ with the target output [H-ML-MH-MLH], the output candidates generated following the above guidelines are listed in Tableau 4.15. Among the five candidates, (a) is the faithful output, and (b) is the target output. The last three candidates (c-e) only differ from the target output in the phrase-final tone. With the initial ranking, the output (c) is optimal since *MLH is higher-ranked, which triggers the demotion of *MLH and the promotion of MAX-LINK(H/ML_).

/H-MLH-MH-MLH/	*MLH	MAX-LINK(M)	MAX-LINK (L/M_H)	MAX-LINK (H/ML_)
a. H-MLH-MH-MLH	**			
b. ✓H-ML-MH-MLH	*→			*
c. ☞☹H-ML-MH-ML				←**
d. H-ML-MH-MH			*	
e. H-ML-MH-H		*		*

Tableau 4.15. Limited types of output candidates of the input /H-MLH-MH-MLH/ with the initial ranking; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

Consider another two-word input /MH-HL/ whose target output is [ML-HL] (MH represents the MH allomorph emerged from T3S). The lone possible output candidate is the faithful output [MH-HL] to which the target output is harmonically-bounded due to the absence of any constraint that could force MH to surface as ML in a non-final position. In other words, when the Tone 3 allomorph MH is selected as the input in the HS context, the output is always an error as shown in Tableau 4.16 (see also Tableau 4.9 in §2.2). The above filtering process only excludes redundant output candidates, and the goal of modeling the Tone 3 change presumably caused by recognizing a different basic allomorph should not be severely affected by reducing the number of output candidates in grammar learning.

/MH-HL/	MAX-LINK
a. ✓ML-HL	*!
b. ☞☹MH-HL	

Tableau 4.16. Surface errors generated from an underlying MH allomorph in the HS context

Finally, unlike the Dutch simulation in which the target output of a lexical item remains identical, the target output of each Tone 3 word may vary probabilistically due to the free variation between [ML] and [MLH] in the phrase-final position. For example, if the learners perceive 35 [ML] tokens and 65 [MLH] tokens of the same Tone 3 word, the chances for the learners to treat each variant as the target output in every production of the Tone 3 word are 35% and 65% respectively. In other words, if this Tone 3 word is produced as [MLH], the chance is 35% for the learners to believe that they produce an output error.

4.3 Lexical factors

Similar lexical factors are used in the following simulations to calculate the SP of each allomorph of Tone 3 words, such as Individual Error Proportion (IEP). Nevertheless, since Morphological Error Proportion (MEP) refers to a consistent morphosyntactic contexts shared by a set of surface allomorphs, which is absent for any set of Tone 3 allomorphs, this lexical factor is assumed to play no role in the simulation based on PSI-OT-GLA. For example, the presence of the full concave allomorphs MLH and the half-tone allomorph ML depends on different phrasal position, and the contrast phrase-final vs. non-final cannot be explained with any consistent morphosyntactic generalization; e.g. an ML allomorph produced in a non-final position might represent noun, and another ML allomorph in the same position might represent a verb. Therefore, IEP will be the lone factor contributing to the calculation SP, which is only involved in the M stage simulation.

The number of output errors generated from an allomorph likewise contributes to the calculation of IEP, but the accumulation of output errors is less straightforward in the Mandarin simulation. In the Dutch simulation, each input is a single word containing one stem, and the selected stem allomorph is necessarily responsible for the produced output error (i.e. the number of output errors increase by one). The inputs in the Mandarin simulation, on the contrary, may have more than one Tone 3 word, and the question is when the optimal output is an error, which Tone 3 allomorphs should be penalized for the error. The current algorithm assumes that the number of output errors of a Tone 3 allomorph accumulates only when the Tone 3 word does not surface as the target output. Two examples are compared in Figure 4.11 below. When the input is /ML-T-MLH/, which surfaces as an output error *[ML-T-ML] due to the unexpected phrase-final reduction, only the phrase-final Tone 3 allomorph MLH is responsible for the error; the phrase-initial ML allomorph surfaces as the target and thus can be exempted from the penalty. Both Tone 3 allomorphs in the input /<u>MH</u>-T-MLH/ are nevertheless penalized since their output does not match the target output. This word-specific evaluation of output error penalties was applied throughout the Mandarin simulations.

Error Count:	+1	+1 +1
Input:	/ML-T-MLH/	/MH-T-MLH/
	↓	↓
Output:	*[ML-T-ML]	*[MH-T-ML]
Target:	[ML-T-MLH]	[ML-T-MLH]

Figure 4.11. Word-specific evaluation of output errors

Finally, it is still assumed that memory decay may erase tonal allomorphs with a relatively low token frequency from the lexicon. Due to a scattered distribution of the learning inputs fed to learners of Generation 1, the decay rate will be conservatively set as 0.0005 per learning cycle. This setting predicts that if a tonal allomorph is perceived around five times for every 10,000 learning cycles, it could completely decay from the memory or is lexically unstable. For Generation 1 learners, the full concave allomorph of 险('danger', 11 tokens per 10,000 cycles), 桶 ('bucket', 8 tokens per 10,000 cycles), 补('mend', 12 tokens per 10,000 cycles), 俩 ('pair', 6 tokens per 10,000 cycles), 产('yield', 3 tokens per 10,000 cycles), and 使 ('make', 12 tokens per 10,000 cycles) may be saliently affected by such an effect and unable to be selected as the input of these Tone 3 words.

4.4 Perception and production in two-stage learning

In the beginning of every learning cycle, the 'adult' speaker will produce a random phrase following the steps in §4.1. In Generation 0, with the very first machine adult speaker', it is assumed that the full concave allomorph (i.e. MLH) of each Tone 3 word has a dominant SP and will always be selected as the input of each Tone 3 word in the input selection process. This configuration attempts to mimic the initial state before the Tone 3 change begins. Since the Generation 0 machine speaker will be assumed to acquire the target grammar (see below), the two tone sandhi processes (T3S and HS) are automatically applied to every output (i.e. learning input) produced by the Generation 0 speaker. For example, the input /MLH-MLH-MLH/ will always be produced as [MH-MH-MLH] via T3S.[51] The output produced by a

[51] The application of T3S in Standard Mandarin varies with different phrase structures. The well-known example is the contrast between [mai^{213} [hau^{24} tɕiou^{213}]] 'buy good wine' vs. [[mai^{24} hau^{24}] tɕiou^{213}] 'have bought wine'; T3S applies to the innermost bracket first, deriving [hau^{24} tɕiou^{213}] and [mai^{24} hau^{213}] respectively, but [mai^{213} [hau^{24} tɕiou^{213}]] does not undergo T3S again without adjacent Tone 3s. For more details, see Chen (2000). Since the randomly generated learning inputs do not bear any semantic meaning, the phrase structure of each learning input cannot be specified, and T3S

Generation n speaker ($n > 0$) is determined by the grammar acquired by the speaker.

The learner of a generation simply perceives a learning input as the target output of each learning cycle and aims at reproducing the target output. In the P stage, the input of the learner's production in a learning cycle is identical to learning input produced by the adult speaker in that learning cycle. For example, if the adult speaker produce [H-HL-MLH], then the input of the learner's production is also /H-HL-MLH/ and the learners aims at generating the faithful output [H-HL-MLH] for the intended adult form. In the M stage, the learner first identifies the tonal labels in the learning input, and the learning input [H-HL-MLH] is recognized as the input /T1-T4-T3/. The learner will proceed to the same input construction process illustrated above in §4.1, and the input is converted into /H-HL-ML/, /H-HL-MH/, or /H-HL-MLH/, depending on the SPs of the Tone 3 word.

4.5 Learning over generations

Unlike the previous simulations which repeated the same learning process with the same learning inputs, the current modeling of the diachronic change, the learning outcome of one generation becomes the input source for the next generation. The learner of Generation 1 receives learning inputs from the Generation 0 speaker to develop a constraint grammar and lexical variables. When the learning process is over, the Generation 1 learner becomes the adult speaker and uses the acquired grammar and the lexical variables to produce learning inputs for the Generation 2 learner. Gradual changes, if any, can thus be observed in the learning outcome of different generations.

In the current work, the learning process was simulated ten times with 10,000 learning cycles in the P stage simulation and 100,000 learning cycles in the M stage simulation in each learning generation. The learning results such as constraint values and SPs were averaged over the ten simulations as the averaged learning outcome for this generation, which was adopted to produce learning inputs for the next learning generation. A total of ten learning generations were simulated to test the modeling of the Tone 3 change in PSI-OT-GLA.

4.6 Simulated elicitation task

By the end of each simulation, a simulated elicitation task was conducted to obtain productions of the Tone 3 words in a phrase-final position for a comparison with

---

is assumed to apply linearly from the leftmost adjacent Tone 3s to the rightmost ones.

the real experimental results. Each of the eighteen Tone 3 words was produced 100 times with the SP of its allomorphs and the constraint values after the 100,000th learning cycle of the M stage simulation. The production results were then averaged over the 10 simulations in the same learning generation for the most representative results at a specific stage of the diachronic development of Tone 3.

4.7 Results

The averaged constraint values at the end of the M stage of each learning generation are plotted in Figure 4.12 to illustrate the grammar changes over ten learning generations. Only the constraints *MLH, MAX-LINK(M), MAX-LINK(L/M_H), and MAX-LINK(H/ML_) are included in the discussion since the constraint value of the three top-ranked markedness constraints OCP-MLH, *NONFINAL-MLH, and *LONGLAPSE remain unchanged at the end of the M stage of each generation. The two proposed unnatural markedness constraints NONFINALDIPPING and FINALCONCAVE in §4.2 have never had a chance to be created due to a significant number of exceptions. The average exception number of the former in each learning generation is 3279 out of 3280, which is far from the creation threshold (i.e. 3280 / log(3280) ≈ 405). The average exception number of the latter is 2899 out of 3280, which is also much higher than the same threshold. Without the first unnatural markedness constraint, the input /MH/ cannot surface as [ML] in the HS context (i.e. non-final position), and without the second one, the inputs /ML/ and /MH/ cannot surface as [MLH] in phrase-final positions. I will return to the influence from the absence of these two unnatural markedness constraints shortly.

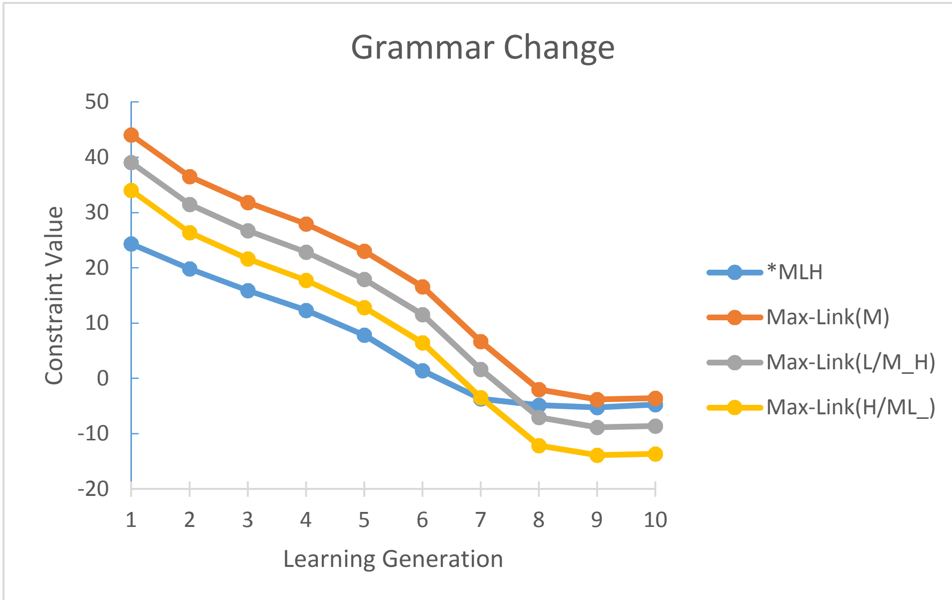


Figure 4.12. Constraint values at the end of the M stage over 10 learning generations

At the end of Generation 1, the constraint ranking MAX-LINK(M) » MAX-LINK(L/M_H) » MAX-LINK(H/ML_) » *MLH is identical to the assumed target ranking of Standard Mandarin: MLH is preserved in phrase-final positions with a lower-ranked *MLH. From Generation 2 to 6, *MLH and MAX-LINK(H/ML_) gradually move closer to each other, which is the process representing the shift of the Tone 3 word target from [MLH] to [ML] as repeated in Tableau 4.17; the number of the /MLH/→[ML] mapping in phrase-final positions is gradually increasing to move the two constraints closer. *MLH is still generally dominated by MAX-LINK due to a greater number of the /MLH/→[MLH] mapping in phrase-final positions. After Generation 7, the /MLH/→[ML] mapping in phrase-final positions finally outnumbers the /MLH/→[MLH] mapping, and a stronger pressure to rank *MLH higher than MAX-LINK emerges. Eventually, *MLH fully dominates MAX-LINK(H/ML_) since Generation 8, prohibiting [MLH] regardless of prosodic positions. The grammar change in the simulation is thus generally identical to the prediction made in §2.2.

T3A /MLH/ Target: ML	*MLH	MAX-LINK
☞✓ML		*
MLH	*!	
T3B /MLH/ Target: MLH	*MLH	MAX-LINK
☞☹ML		*
✓MLH	*!	

Tableau 4.17. The ranking *MLH » MAX-LINK required by the mapping MLH→ML

Two Tone 3 words are selected from each RPU family to illustrate the representative SP variation of the [MLH] allomorph in each RPU family in Figure 4.13: 礼 [li^{213}] with a RPU of 0.568 and 远 [yãn^{213}] with a RPU of 0.337 from the High RPU family, 组 [tsu^{213}] with a RPU of 0.235 and 桶 [t^{h}uõŋ^{213}] with a RPU of 0.188 from the Medium RPU family, and 请 [tɕ^{h}ĩŋ^{213}] with a RPU of 0.027 and 指 [zɨ^{213}] with a RPU of 0.015 from the Low RPU family. Before Generation 7, which is the boundary of the complete grammar change as indicated above, there are in general three different paths of SP changes: For Tone 3 words from the High RPU family, the SP of thei [MLH] allomorph remains fully dominant. For those from the Medium RPU family, the same SP slightly decreases over time and does not change drastically until the complete grammar change after Generation 7. Finally, the SP of the [MLH] allomorph quickly plummets and reaches the floor after 3 or 4 generations for the Tone 3 words from the Low RPU family. Recall the prediction in that Tone 3 words which occur more frequently in phrase-final positions tend to have a higher SP for their [MLH] allomorph. This developmental difference is captured in the simulated results.

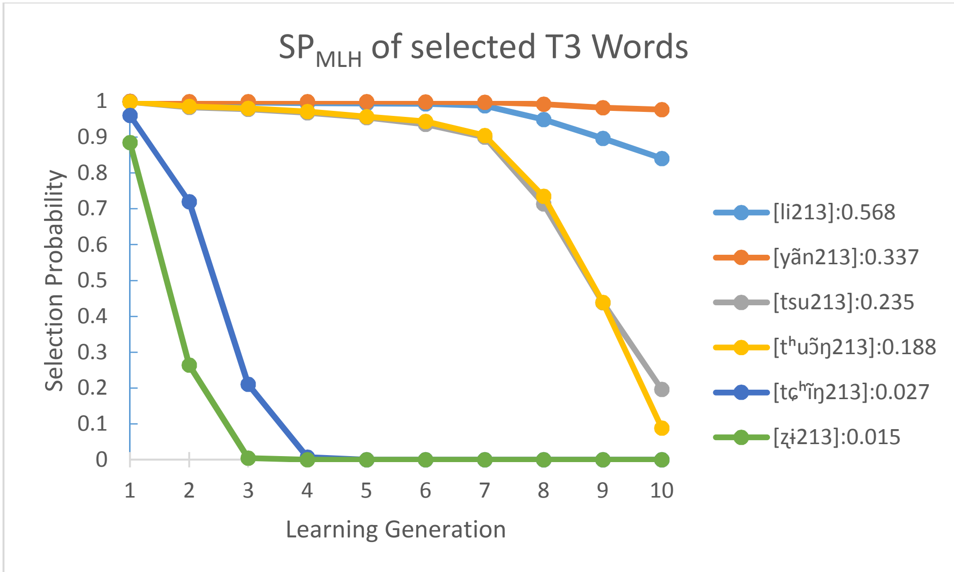


Figure 4.13. SP$_{MLH}$ of the six selected Tone 3 words over 10 learning generations

It is also important to note that after SP$_{MLH}$ drops for the Tone 3 words, it is always the ML allomorphs taking over the dominant status in the input selection. Along the ten learning generations, SP$_{MH}$ is never higher than 0.05. The reason repeated here is twofold. First, the MH allomorph has a lower token frequency because of its limited occurrence in the T3S context. The MH allomorph therefore does not have the same frequency advantage as the ML allomorph. Second, the MH allomorph can only surface as the target output in the T3S context with a natural grammar, and the two unnatural markedness constraints are not created for deriving the target output in either the HS context (i.e. /<u>MH</u>/→[ML]) or a phrase-final position (i.e. /<u>MH</u>/→[MLH]).

With the SPs and acquired grammar at the end of the M stage simulation of each generation, the machine learner was required to produce the eighteen target Tone 3 words in a phrase-final position as in the real experiment in §3. The MLH proportions of the same six selected words are illustrated in Figure 4.14. Recall that in PSI-OT-GLA the input can be either /MLH/ and /ML/, which can surface faithfully as [MLH] and [ML], but the MLH proportions are not merely determined by the SP$_{MLH}$ and SP$_{ML}$ of each Tone 3 word: There is a chance for *MLH to outrank MAX-LINK, and the input /MLH/ thus surfaces as the half-tone variant [ML]. Therefore, the MLH proportion of each Tone 3 word is always lower than SP$_{MLH}$ of the Tone 3 word.

Before the significant grammar change (i.e. Generation 7), the probability of producing a Tone 3 word as [MLH] in phrase-final positions correlates with SP_{MLH} of the Tone 3 word. The two Tone 3 words from the High RPU family are almost always realized as [MLH] with a dominant SP_{MLH}. The slightly falling MLH proportions for the High RPU family before Generation 7 are not the consequence of the SP change (see Figure 4.13) but the gradual grammar change that moves *MLH and MAX-LINK closer over time and allows an input /MLH/ to surface as [ML] as noted above. The two Tone 3 words from the Medium RPU family also have a chance higher than 90% to be realized as [MLH]. Finally, with a drastic decrease in SP_{MLH} of the two Tone 3 words from the Low ratio family, the ML allomorph is more likely to be selected as the input, and thus the two Tone 3 words are less likely to be realized as [MLH]. When SP_{MLH} drops to almost zero in Generation 4, the two Tone 3 words can never surface as [MLH] even if the grammar change has not finalized.

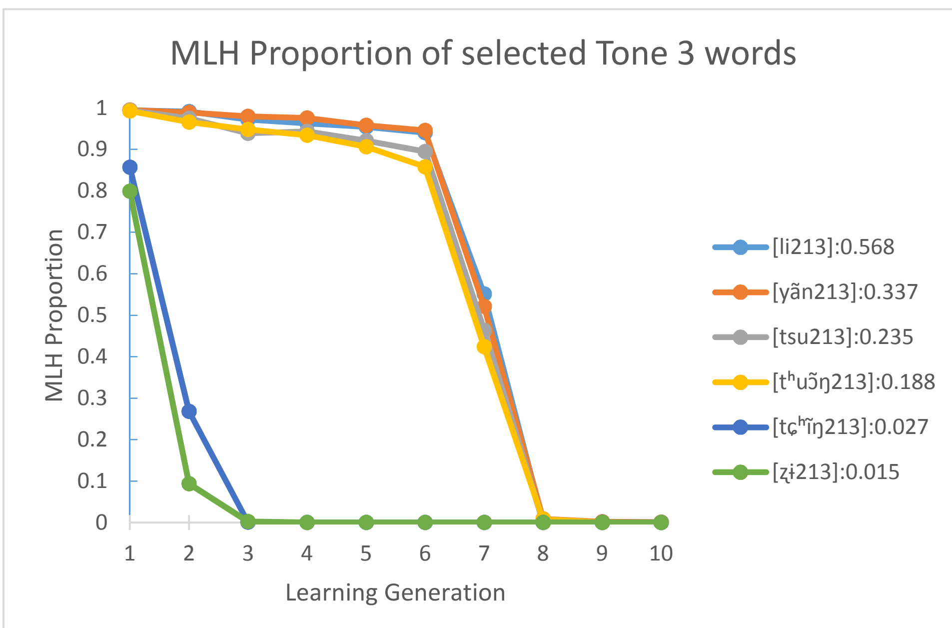


Figure 4.14. MLH proportion of the six selected Tone 3 words in the simulated elicitation task over ten learning generations

Finally, let us consider the similarity between the simulation and experimental results. Since there are ten learning generations, the first step is to decide which generation may have the concave percentages most similar to those in the experimental result. To this end, a series of linear regression tests were used with the concave percentages as the dependent variable and Ratio as the sole independent variable. The last three learning generations are excluded from the tests since all Tone 3 words are realized as ML, which is clearly not the case demonstrated by the experimental results. The six Tone 3 words from the Low RPU family are also excluded as outliers from each test since they are also completely realized as [ML] after Generation 2 as opposed to the real production data. The analysis results in Table 4.7 show that the predictor Ratio is marginally significant only in Generation 7.

Generation	1	2	3	4	5	6	7
t-value	1.02	1.29	1.13	1.1	1.25	1.38	1.86
p-value	0.333	0.23	0.28	0.299	0.243	0.201	**0.096**

Table 4.7. P-values of the sole independent variable Ratio in linear regression tests using concave percentages as the dependent variable

The comparison of concave percentages of the twelve Tone 3 words from the High and Medium RPU family is illustrated in Figure 4.15 below. Out of the twelve Tone 3 words in the comparison, six words have a difference within 5% between the experimental and simulated data, and another two words have a gap lower than 10%. Unfortunately for the Tone 3 word 喜 [ɕi^{213}] 'happy', like the Tone 3 words of the Low ratio family, PSI-OT-GLA predicts that the output [MLH] is already impossible in Generation 7. Otherwise, the simulated result in Generation 7 generally matches the experimental results in terms of individual concave percentages and the gradually decreasing MLH production rates from higher ratio Tone 3 words to lower ones.

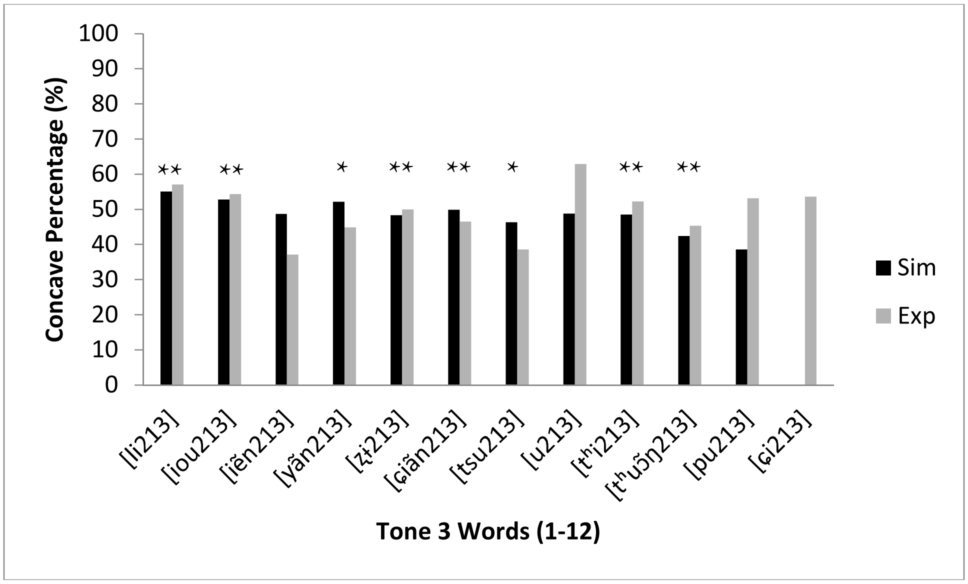


Figure 4.15. Comparison of concave percentages of the 12 Tone 3 words from the High and Medium ratio family between the simulated and experimental results; * difference ≤ 10%, ** difference ≤ 5%

4.8 Discussion

The similarity between the experimental and simulated results in the above sections suggest that PSI-OT-GLA is at least on the right track to predict historical changes triggered by gradually selecting different allomorphs over time, which eventually lead to the ultimate grammar change. That being said, I have no intention to ignore the disparity that occurs between the results of the low ratio Tone 3 words in the experiment and simulation. In particular, even if a stronger Ratio effect can be observed with a more natural experiment setting, a sharp contrast between high and low ratio Tone 3 word productions is still very unlikely, i.e. some variations in high RPU Tone 3 words, but absolutely no variation in low RPU Tone 3 words as predicted in PSI-OT-GLA.

Recall that the lack of variations in low RPU Tone 3 words before Generation 7 in the simulation lies in the extremely low token frequency of their MLH allomorph, which in turn results in a low SP_{MLH}. Furthermore, since MEP was excluded from the simulation, the SP_{MLH} of each Tone 3 word is calculated independently with only the corresponding IEP_{MLH}; although high RPU Tone 3 words prefer the MLH allomorph to be the input, such a generalization cannot spread through the lexicon to raise the SP_{MLH} of low RPU Tone 3 words.

This tonal change is nevertheless a change of the entire phonological category (i.e. Tone 3), and one can thus imagine an across-the-board perseverance or reduction of the MLH realization for all Tone 3 words. That is, it is possible to assume a lexical structure akin to an exemplar network in which all Tone 3 words are associated together through this 'Tone 3' label (Figure 4.16). This type of lexical association is also consistent with the findings that although in general tonal priming effect is absent (e.g. Chen & Chen 2002), it is used to rule out tonal mismatch competitors or non-words (i.e. items with a different tonal 'label') during an early phase of lexical access (e.g. Lee 2007, Shuai et al. 2012). In light of this assumption, the production of a low RPU Tone 3 word may be affected by the production of a high RPU Tone 3 words via their lexical association between them, which can be modeled with Output-Output (OO) correspondence. For example, the production of a low RPU Tone 3 word 指 [ʐɨ^{213}] 'point' is required to be leveled to the full concave allomorph of a high RPU Tone 3 word 止 [ʐɨ^{213}] 'stop' by an OO constraint and thus surface with a full concave tone as well.

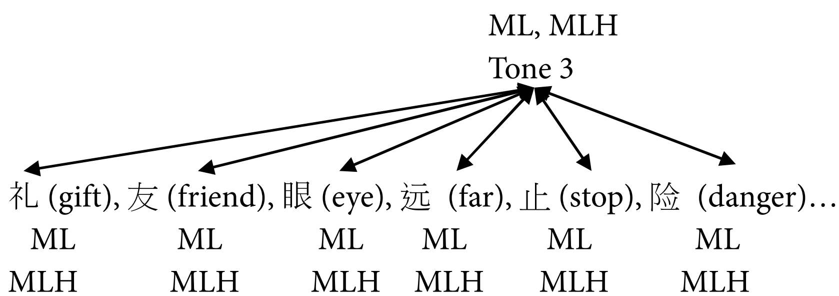


Figure 4.16. Hypothetical lexical structure of Tone 3 words with a 'Tone 3' label

As the full concave variant is still considered the standard pronunciation of Tone 3, the pressure that requires a Tone 3 word output to be leveled to a full concave paradigm might be higher. The diachronic development may thus proceed slowly over generations, and the MLH proportion of low RPU Tone 3 words does not dramatically diverge from that of high and medium RPU ones. Paradigmatic uniformity between Tone 3 words is not available in PSI-OT-GLA due to the exclusion of OO constraints, but I will return to this issue in Ch. 6 in the discussion of a possible OO model.

5. Local summary

In this chapter, PSI-OT-GLA has been tested with the hypothesis that the Tone 3 change from a full concave tone to a half-tone in Standard Mandarin is driven by a gradual shift from the full concave allomorph MLH to the half-tone ML allomorph. The experiment and simulated results both support this hypothesis and demonstrated the effect of different frequency distributions on the changing progress of individual Tone 3 words. Importantly, the contour simplification process in Standard Mandarin and the overgeneralization patterns produced by Dutch-learning children in Ch. 3 are linked in the similar way: Both are triggered by selecting different surface allomorphs as inputs and thus cannot be accounted for without positing a rich lexicon that stores surface allomorphs and the lexical information crucial to the determination of basic allomorphs. Unlike the Dutch case, however, an adult-like lexical generalization and constraint grammar is not likely to be acquired in Standard Mandarin as predicted in the simulation, and the shift in input selection is thus expected to be permanent as the simulation continues. It is not an unusual case that diachronic sound changes can frequently be discovered in child phonological acquisition as they sprout from the same 'seeds' (e.g. same articulatory or perceptual motivation; see Ohala 1975, 1978; Ohala & Lorentz 1977), and PSI-OT-GLA provides another piece of evidence for such an association in terms of morphophonological acquisition, which stresses the role of a rich lexicon in both synchronic and diachronic language development.

Chapter 5

Probabilistic Selection of Input and stem-final obstruent variations in Korean

This chapter continues to focus on relexicalization triggered by a frequency bias in the selection process of input with stem-final obstruent variations in Korean suffixed noun forms, which has been briefly summarized in §3.2 of Ch. 1. Unlike the relexicalization process in Standard Mandarin which triggers a single grammar change, relexicalization in Korean generates stem-final obstruent variations by reversing the original many-to-one positional neutralization process, and PSI-OT-GLA was implemented to model this diachronic development to see whether the algorithm can be further verified by capturing this complicated development. In §1, a conventional OT analysis is offered for illustrating the morphophonemic alternation via coda neutralization and assibilatory affrication, which create complex allomorphic patterns. The details of stem-final obstruent variations in Korean suffixed noun forms collected in Jun & Lee's (2007) experimental study will be given in §2, followed by a discussion of possible triggers of the changes and predictions made with PSI-OT-GLA in §3. The section §4 consists of both the simulation designs and the comparison between the simulation and experimental results in Jun & Lee (2007).

1. Standard Korean phonology of noun stem-final alternations

The chapter beings by constructing the Korean OT phonology of noun stem-final alternations according to the standard pronunciation of thirteen noun stems in different suffixal contexts listed in Table 5.1, including no suffixation (bare form), topic (/-ɨn/), nominative (/-i/), accusative (/-ɨl/), directive/instrumental (/-ɨlo/), and locative/dative (/-e/).

Stem	Bare	Topic	Nominative	Accusative	Directive/ Instrumental	Locative/ Dative
sickle	nat̚	nas-ɨn	nas-i	nas-ɨl	nas-ɨlo	nas-e
clothes	ot̚	os-ɨn	os-i	os-ɨl	os-ɨlo	os-e
day	nat̚	natʃ-ɨn	natʃ-i	natʃ-ɨl	natʃ-ɨlo	natʃ-e
breast	tʃət̚	tʃətʃ-ɨn	tʃətʃ-i	tʃətʃ-ɨl	tʃətʃ-ɨlo	tʃətʃ-e
flower	k’ot̚	k’otʃʰ-ɨn	k’otʃʰ-i	k’otʃʰ-ɨl	k’otʃʰ-ɨlo	k’otʃʰ-e
face	nat̚	natʃʰ-ɨn	natʃʰ-i	natʃʰ-ɨl	natʃʰ-ɨlo	natʃʰ-e
field	pat̚	patʰ-ɨn	patʃʰ-i	patʰ-ɨl	patʰ-ɨlo	patʰ-e
red bean	pʰat̚	pʰatʰ-ɨn	pʰatʃʰ-i	pʰatʰ-ɨl	pʰatʰ-ɨlo	pʰatʰ-e
kitchen	puək̚	puəkʰ-ɨn	puəkʰ-i	puəkʰ-ɨl	puəkʰ-ɨlo	puəkʰ-e
outside	pak̚	pak’-ɨn	pak’-i	pak’-ɨl	pak’-ɨlo	pak’-e
wall	pjək̚	pjək-ɨn	pjək-i	pjək-ɨl	pjək-ɨlo	pjək-e
leaf	ip̚	ipʰ-ɨn	ipʰ-i	ipʰ-ɨl	ipʰ-ɨlo	ipʰ-e
rice	pap̚	pap-ɨn	pap-i	pap-ɨl	pap-ɨlo	pap-e

Table 5.1. Standard pronunciation of thirteen noun stems in different morphosyntactic contexts

Two crucial phonological processes in standard Korean OT phonology are discussed in this section, including (i) obstruent coda neutralization shown by the bare forms in Table 5.1, which will be analyzed in §1.1 and (ii) assibilatory affrication which turns plain stops into affricates before a high front vowel demonstrated by nominative forms in Table 5.1, which will be discussed in §1.2. Both alternations create multiple surface allomorphs that might complicate the learners' identification of correct URs and also a set of constraints whose ranking may change over time if learners fail to recognize correct URs (see §2 and §3).

1.1 Obstruent coda neutralization in bare forms

Korean has a rich obstruent inventory due to a three-way laryngeal contrast (plain, aspirated, and tense) as in (1). Regardless of the laryngeal specification, however, obstruents are neutralized to a homorganic plain unreleased stop in a syllable coda position as revealed by the morphophonemic alternations in (2).[52]

[52] Kim & Jongman (1996) concluded in a phonetic investigation that 83% of word-final stops are followed by a short burst, which may be evidence against defining word-final stops as strictly unreleased.

(1) Coronal obstruents in Korean
plain: /p, t, k, tʃ, s/
aspirated: /pʰ, tʰ, kʰ, tʃʰ/
tense: /p’, t’, k’, tʃ’, s’/[53]

(2) Korean coronal coda neutralization (from Martin 1992 and Jun & Lee 2007)

e.	[pat̚]	‘field’	[patʰ-ɨl]	‘field (acc.)’	Stem UR: /patʰ/
f.	[tʃʌt̚]	‘milk’	[tʃʌdʒ-ɨl]	‘milk (acc.)’	Stem UR: /tʃʌtʃ/[54]
g.	[k’ot̚]	‘flower’	[k’otʃʰ-ɨl]	‘flower (acc.)’	Stem UR: /k’otʃʰ/
h.	[ot̚]	‘clothing’	[os-ɨl]	‘clothing (acc.)’	Stem UR: /os/
i.	[ip̚]	‘leaf’	[ipʰ-ɨl]	‘leaf (acc.)’	Stem UR: /ipʰ/
j.	[puək̚]	‘kitchen’	[puəkʰ-ɨl]	‘kitchen (acc.)’	Stem UR: /puəkʰ/

In a standard generative analysis, the stem input is assumed to be its accusative allomorph and the stem-final neutralization can be considered a coda neutralization process. These native phonotactics are also strictly abided by for English loanwords in Korean. For example, Jun (2002), Kang (2003), and many others found that word-final stops in English words are adapted either as plain unreleased stops or as aspirated stops followed by an inserted vowel in (3). Vowel insertion usually emerges from a long word-final release in the source English words, which is considered a possible vowel hallucination effect (see Dupoux et al. 1999, Boersma & Hamann 2009, Kang 2003). The vowel /ɨ/ is often devoiced in Korean native words particularly in a non-initial open syllable (Kim et al. 1993), which is acoustically similar to a long release after a stop. Korean listeners thus tend to compensate for this acoustic information loss during perception, and English words with a salient

[53] The tense obstruents are transcribed as ejectives as in Jun (2010), but they in fact contrast with plain and aspirated obstruents in a far more complex way. Articulatorily, tense obstruents involve various laryngeal processes such as stiffening of the vocal folds and lowering of the glottis (e.g. Kagaya 1974). Acoustically, Cho et al. (2002) demonstrated a shorter VOT, lower burst energy, and higher f0 in the following vowel for tense stops and a higher centroid frequency and f0 in the following vowel for tense fricatives. Readers are referred to Halle & Stevens (1971), Keating (1984), Kim & Duanmu (2004), and many others for different accounts that develop the phonological representation of these tense obstruents from the above properties.

[54] While obstruent voicing contrast is not phonemic in Korean as shown in (1), there is an allophonic rule that changes voiceless plain stops into voiced stops intervocalically (Jun 1993, Silva 1992). Jun (1993) specifically argues that plain affricates may not always undergo the process, and fricatives are rarely produced with voicing.

word-final stop release are frequently borrowed with /ɨ/-insertion.

(3) Vowel insertion after released word-final stop in Korean English loanwords (Kang 2003)

mint → [min.tʰɨ]

hit → [hi.tʰɨ]

The OT analysis of the neutralization process is illustrated in Tableau 5.1, in which tense obstruents are assumed to be [+tense] and plain obstruents to be [-tense]. Positional markedness constraints (4a), (4b), and (4c) outrank general faithfulness constraints (4d), (4e), and (4f) changing the tense and aspirated coronal consonants into a plain stop (e.g. Boersma & Hamann 2009). This grammar is an example showing different contrast numbers allowed in different syllable positions (e.g. Beckman 1999; Lombardi 2001). Whether perceptual asymmetry (e.g. Steriade 2001b) is the root of this asymmetry requires further research.

(4) OT constraints for coda neutralization in Korean

a. *+strid/_]σ: Syllable-final stridents (include /s/ and /tʃ/) are prohibited.

b. *+asp/_]σ: Syllable-final aspirated consonants (include /pʰ/, /tʰ/, /kʰ/, /tʃʰ/) are prohibited.

c. *+tense/_]σ: Syllable-final tense consonants (include /p'/, /t'/, /k'/, /s'/) are prohibited.

d. IDENT(strid)-IO: Input [strid] specification must be preserved in the output.

e. IDENT(asp)-IO: Input [asp] specification must be preserved in the output.

f. IDENT(tense)-IO: Input [tense] specification must be preserved in the output.

/CVs/	*+strid/_]σ	*+asp/_]σ	*+tense/_]σ	IDENT (strid)	IDENT (asp)	IDENT (tense)
☞CVt̚				*		
CVs	*!					
/CVp^{h}/						
☞CVp̚					*	
CVp^{h}		*!				
/CVp'/						
☞CVp̚						*
CVp'			*!			

Tableau 5.1. Coda contrast neutralization in Korean; /CVs/ represents all inputs with a strident coda, /CVp^{h}/ represents all inputs with an aspirated coda, and /CVp'/ represents all inputs with a tense coda

On the other hand, onset coronal contrasts are correctly generated by ranking general markedness constraints (5a), (5b), and (5c) lower than faithfulness constraints as in Tableau 5.2.

(5) General markedness constraints forbidding strident, aspirated, and tense consonants

a. *+strid: Stridents (include /s/ and /tʃ/) are prohibited.

b. *+asp: Aspirated consonants (include /p^{h}/, /t^{h}/, /k^{h}/, /tʃ^{h}/) are prohibited.

c. *+tense: Tense consonants (include /p'/, /t'/, /k'/, /s'/) are prohibited.

/sV/	IDENT (strid)	IDENT (asp)	IDENT (tense)	*+strid	*+asp	*+tense
☞sV				*		
tV	*!					
/p^hV/						
☞p^hV					*	
pV		*!				
/p’V/						
☞p’V						*
pV			*!			

Tableau 5.2. Onset contrast preservation in Korean; /sV/ represents all inputs with a strident onset, /p^hV/ represents all inputs with an aspirated onset, and /p’V/ represents all inputs with a tense onset

1.2 Assibilatory affrication

The Korean morphophonology is further complicated with the assibilatory affrication process which turns stem-final coronal stops into palatal-alveolar affricates before suffixes starting with /-(h)i/ but not before any /-ɨ/-initial suffix as shown in (6).

(6) Restricted assibilatory affrication in Korean

a. Before /-i/

/mat+i/	→	[matʃ-i]	‘first child (nom.)’
/pat^h+i/	→	[$patʃ^h$-i]	‘field (nom.)’

b. Before /-ɨ/

/pat^h+ɨn/	→	[pat^h-ɨn]	‘field (top.)’
/p^hat^h+ɨl/	→	[p^hat^h-ɨl]	‘red bean (top.)’

An OT analysis accounting for the occurrence (and prohibition) of assibilatory affrication requires two markedness constraints *Ti and *Tɨ in (7), which are phonetically natural since the tongue gesture of a high vowel creates a narrow passage that results in frication after the release of the coronal closure (Kim 2001; Kirchner 2001; cf. Hall et al. 2006).[55] A similar process exists in Québecois French

[55] Assibilatory affrication before a high central vowel is rare, perhaps because of the typological rarity of high central vowels which prevents synchronic alternations or diachronic changes to be observed. Coarticulation may also result in the low frequency of high central vowels following /tʃ/ as high

(e.g. Papen 1998) and Japanese (e.g. Ito & Junko 1995). The markedness constraint *Ti must outrank IDENT(del rel)-IO in (8), which in turn outranks *Tɨ to account for the patterns in (6) (see Tableau 5.3). This sub-ranking *Ti » IDENT(del rel) » *Tɨ is typologically common as both cross-linguistic surveys (e.g. Kochetov 2011) and phonetic investigations indicate that assibilatory affrication is mostly triggered by high front vocoids /i/ and /j/ (e.g. Hall et al. 2006, Kim 2001).

(7) Constraints for Korean assibilation

a. *Ti: A sequence of coronal stop + high front vowel is prohibited (see also Hall & Hamann 2006).

b. *Tɨ: A sequence of coronal stop + high central vowel is prohibited.

(8) IDENT(del rel)-IO: Input [delayed release] specification must be preserved in the output.

/mat+i/	*Ti	IDENT(del rel)	*Tɨ
☞ma.tʃ-i		*	
ma.t-i	*!		
/patʰ+ɨn/			
☞pa.tʰ-ɨn			*
pa.tʃʰ-ɨn		*!	

Tableau 5.3. OT analysis of Korean assibilatory affrication

Another faithfulness constraint IDENT(cont)-IO in (9a) is also required to dominate LAZY in (9b) to avoid satisfying *Ti by changing the marked structure /t+i/ to [si] as in Tableau 5.4, which is possible as observed in the lenition process in the inherently affricated context (e.g. before high front vocoid) in Ancient Greek, Nez Perce, and Turkana (Kirchner 2001; see also Kaplan 2010).

central vowels are easily fronted in this context; in Lahu, the coarticulation effect was phonologized as a phonotactic *tʃɨ (Flemming 2003). Since high back vocoid /ɯ/ (Japanese) and /w/ (Lomongo; Kenstowicz & Kisserberth 1977) can trigger the assibilatory affrication of coronal stops, high vowels, including central ones, should be in general better triggers of assibilatory assimilation than mid and low vowels (e.g. Hall & Hamann 2006, Telfer 2006). The markedness constraint *Tɨ and are thus phonetically motivated.

(9) Constraints for a possible lenition process from /tʃ/ to /s/

a. IDENT(cont)-IO: Input [continuant] specification must be preserved in the output.

b. LAZY: Affricate (****) > Strident Fricative (***) >Stop (**) >Non-strident Fricative (*)

/mat+i/	*Ti	IDENT(cont)	IDENT(del rel)	LAZY
☞ [ma.tʃ-i]			*	****
[ma.s-i]		*!		***
[ma.t-i]	*!			**

Tableau 5.4. Preserving affricates with a top-ranking IDENT(cont)-IO

Finally, assibilatory affrication in Korean only applies in a morphologically derived environment and thus obeys the Strict Cycle Condition (SCC; Kiparsky 1982b); the word /t^h^i/ 'blemish' surfaces faithfully as [tʰi], rather than *[tʃʰi]. SCC in this assibilatory affrication case can be modeled by different constraint models, including local conjunction constraints against marked structures and misalignments between stem and syllable boundaries simultaneously (e.g. *Ti & ANCHOR-R(stem, σ); Lubowicz 2002), and a faithfulness constraint forbidding homomorphemic feature spreading (e.g. HOMCONSISTENCY; Horwood 2006). For the reason of simplicity, I adopt Beckman's (1999) IDENT(σ_1)-IO in (10) to preserve any alternation within a stem since the Korean stems analyzed in this chapter are all monosyllabic; i.e. the consonant in the first syllable is always in a non-morphologically derived environment. The hypothetical input /tit+i/ with a top-ranked IDENT(σ_1)-IO in Tableau 5.5 is sufficient to generate the intended output [ti.tʃ-i] with a restricted application of assibilatory affrication.

(10) IDENT(σ_1)-IO: Segmental features in the first syllable must be preserved in the output.

/tit+i/	IDENT(σ_1)-IO	*Ti	IDENT-IO
☞ [ti.tʃ-i]		*	*
[ti.t-i]		**!	
[tʃi.tʃ-i]	*!		

Tableau 5.5. Non-derived environment blocking of assibilatory affrication

1.3 Summary: Constraint ranking for Standard Korean suffixed noun forms

An additional sub-ranking is required for the different constraints proposed in the above two sections: *Ti must dominate IDENT(strid)-IO as well for assibilatory affrication as in Tableau 5.6. Considering the sub-rankings altogether, the full constraint ranking of standard Korean pronunciations can be summarized as Figure 5.1.

/mat+i/	*Ti	IDENT(strid)
☞ [ma.tʃ-i]		*
[ma.t-i]	*!	

Tableau 5.6. OT analysis of assibilatory affrication with the sub-ranking *Ti » IDENT(strid)-IO

{*+strid/_]σ *+asp/_]σ *+tense/_]σ IDENT(σ_1)-IO}

|

*Ti

|

{IDENT(tense)-IO IDENT(asp)-IO IDENT(strid)-IO IDENT(cont)-IO}

|

IDENT(del rel)-IO

|

{*+asp *+strid *+tense LAZY *Tɨ}

Figure 5.1. Constraint ranking for coda neutralization and assibilatory affrication in Korean

2. Stem-final obstruent variations in Korean suffixed noun forms

Phonologists have generally agreed on the above traditional neutralization analysis with a single UR for each stem, which has only faced infrequent challenges (e.g. Albright 2002, 2008; Jun 2010). Surface variations of suffixed noun forms documented since Martin (1992), however, suggest that the morphophonemic alternation in Korean has been reanalyzed, and an alternative account is in need.[56]

[56] See also Albright & Kang (in press) for a discussion and analysis of the innovation of Korean verb

In (11), it is shown that the bare forms have remained unchanged over time, but several surface variants have emerged for their accusative forms. The variants [pas-ɨl] and [k’os-ɨl] demonstrates the extension of the [t]~[s] alternation from the [ot̚]~[os-ɨl] pair, and the variants [patʃʰ-ɨl] and [tʃʌdʒ-ɨl] exhibit the extension of the [t]~[tʃʰ] alternation from the [k’ot̚]~[k’otʃʰ-ɨl] pair.

(11) Extended alternation in Korean (Martin 1992)

e.	[pat̚]	‘field’	[patʰ-ɨl], [patʃʰ-ɨl], [pas-ɨl]	‘field (acc.)’
f.	[tʃʌt̚]	‘milk’	[tʃʌs-ɨl], [tʃʌdʒ-ɨl]	‘milk (acc.)’
g.	[k’ot̚]	‘flower’	[k’otʃʰ-ɨl], [k’os-ɨl]	‘flower (acc.)’
h.	[ot̚]	‘clothing’	[os-ɨl]	‘clothing (acc.)’[57]

To further investigate the variation patterns, Jun & Lee (2007) conducted elicitation experiments collecting the production of stem-final obstruents before different suffixes in both native words and English loanwords from ten native Korean speakers. The results of their Experiment I (native words) listed with the UR of each stem in a standard analysis in Table 5.2, where numbers stand for the frequency of each stem-final obstruent used in different suffixal contexts.[58]

paradigms.

[57] The surface variation of ‘clothing (acc.)’ was not documented by Martin (1992), but a token of [otʃʰ-ɨl] was recorded in Jun & Lee’s (2007) elicitation experiment (Exp I). The minor difference can be either incidental or a continuous diachronic change from 1992 to 2007.

[58] Jun & Lee represented the affricates as /č/ and /čʰ/, which are replaced with /tʃ/ and /tʃʰ/ respectively in Table 5.2.

Stem-final C	Stem	nominative *-i*					accusative *-ɨl*					locative/dative *-e*				
		s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
s	/nas/ 'sickle'	10					10					10				
	/os/ 'clothes'	10					9	1				10				
tʃ	/natʃ/ 'day'	1			9		4			6					10	
	/tʃətʃ/ 'breast'	5			5		9			1		None[59]				
tʃʰ	/k'otʃʰ/ 'flower'	2	8				4	6					9	1		
	/natʃʰ/ 'face'	3	7				4	6				5	6			
tʰ	/patʰ/ 'field'	1	9				2	8					1	9		
	/pʰatʰ/ 'red bean'	5	6				6	3	1			5	1	4		1

Table 5.2. Token numbers of stem-final coronal obstruents used in suffixed noun forms (Jun & Lee 2007, Experiment I)

Among the eight noun stems with a stem-final coronal obstruent in their suffixed forms, only the two stem URs with a stem-final /s/ (i.e. /nas/ and /os/) were almost always produced as the standard pronunciation across different suffixal contexts. The other six stems were rather produced with both inter-context and inter-stem variations: Stem-final /tʃʰ/ in /k'otʃʰ/ and /natʃʰ/ was more likely to surface faithfully before the nominative suffix [-i], whereas in the accusative context there was a weak bias toward the realization with a stem-final [s]. The stems /patʰ/ and /pʰatʰ/, while having the same stem-final obstruent in their UR, generated different distributional patterns; the former was frequently produced with a stem-final [tʃʰ] except before the locative/dative suffix [-e], but for the latter stem-final [s] variants seemed more productive. Overall the paucity of stem-final [t] variants is true for each stem in each suffixal context. Due to a small number of participants,

[59] According to Jun & Lee, the tokens of [tʃətʃ-e] 'breast (loc.)' were excluded due to an experimenter's mistake.

the results above might not be a complete picture of surface variations in Korean speakers' production, but the variation tendencies are generally consistent with the well-formedness judgment survey results in Jun (2010): If a suffixed form varies at the surface level, a stem-final [s] is the most favorable target and a stem-final [t] is the least favorable one.

Stem-final variations also occurred in the suffixed forms whose stem has a non-coronal stop in the same experiment as shown in Table 5.3. The intriguing pattern of variation occurred with the stem-final /kʰ/ in the UR /puəkʰ/, which was frequently de-aspirated on the surface before almost every suffix (e.g. [puək-i]). The bilabial stop in the UR /ipʰ/ was aspirated more consistently but could be also de-aspirated before the accusative suffix [-ɨl] (e.g. [ip-ɨl]). Jun (2010) found a more extreme pattern in a well-formedness judgment survey that the rating for the de-aspirated bilabial and velar stop was much higher than the aspirated standard pronunciation in the topic context (i.e. followed by the topic marker [-ɨn]).

Stem-final C	Stem	nominative *-i*			accusative *-ɨl*			locative/dative *-e*		
		plain	asp	tense	plain	asp	tense	plain	asp	tense
kʰ	/puəkʰ/ 'kitchen'	8	2		9	1		6	4	
k'	/pak' / 'outside'			10			10		1	9
k	/pjək/ 'wall'	10			10			10		
pʰ	/ipʰ/ 'leaf'		10		3	7			10	
p	/pap/ 'rice'	9[60]			10			10		

Table 5.3. Stem-final bilabial/velar obstruent variations in Jun & Lee's Experiment I

Previous studies, including Albright (2002), Jun (2010), and Jun & Lee (2007), view the above variations as a consequence of lexical reanalysis. That is, the 'basic' form of Korean nouns has been generally shifted to the bare allomorph (e.g. /nat/ for 'sickle'; cf. §3.2), due to its significantly higher token frequency. Suffixed forms are thus constructed with the bare paradigms in contemporary Korean (e.g. /nat+i/

[60] One output token *[pap-si] was excluded as a likely production error by Jun & Lee.

for 'sickle (nom.)'). These studies also adopt the same approach in which the obstruent derived from an underlying stem-final [t] in a suffixed form depends on the type frequency of obstruents in the stem-final position of suffixed forms. Since stem-final /s/ has the highest type frequency (393 out of 878 suffixed nouns based on Jun's corpus count), the fricative becomes the most prevalent surface variation pattern as shown above in Table 5.2. Other than the fricative, Jun (2010) proposes a preference hierarchy of the coronal variations based on the same frequency count: [s] (393/878) » [tʰ] (253/878) » [tʃʰ] (200/878) » [tʃ] (32/878) » [t] (0/878).[61] The suffixed form variants with a stem-final [tʰ] or [tʃʰ] are indeed attested in Table 5.3, although occurring less frequently than stem-final [s] variants. The utter absence of a suffixed noun with a stem-final [t] is also consistent with the rarity of suffixed form variants with a stem-final [t] (cf. Korean-learning children in Do (2012); see also §4.7). Likewise, the plain labial/velar stop has the highest type frequency in the stem-final position of suffixed nouns (2193 out 2315 for a stem-final /p/ and 7501 out of 7537 for a stem-final /k/), which could give rise to the devoicing variations in Table 5.3. I will demonstrate how these predictions based on type frequencies are also possible in PSI-OT-GLA in §3 and return to discuss the drawbacks of the pure type frequency account at the end of §4.

3. A PSI-based account

This section illustrates different predictions made with the current PSI account. Sections §3.1 and §3.2 presume the lexical change that initiated the stem-final variations in Korean nouns as in previous studies: The basic allomorph has been shifted from the allomorphs in a morphologically complex context to the bare form allomorphs. Top-ranked unnatural markedness constraints are also invented by learners to derive suffixed nouns from the inputs with bare form allomorphs. As conflicting sub-rankings of faithfulness constraints are required to generate correct outputs, the most dominant sub-ranking determines which stem-final alternation is extended to other suffixed nouns. A different view is taken in §3.3 to explain part of the variation patterns: For some noun stems, the basic allomorph is not their bare form allomorph but non-bare form allomorphs (e.g. locative/dative) since some

[61] Jun (2010) in fact decomposed the frequency distribution into suffix-specific distributions to predict the preferred stem-final variants in different suffixed forms, which will be ignored in this chapter.

noun stems might be in fact produced in non-bare form contexts more frequently. In particular, locative/dative allomorphs may have a high token frequency as well because the locative/dative suffix is never dropped even in conversational speech. Finally, §3.4 consists of predictions of how the natural constraint grammar may evolve after the surface variations become learning inputs for the learners of the next generation.

3.1 Against stem-final /t/ in suffixed nouns

PSI-OT-GLA incorporates token frequency as the base of the probability calculation, so it should not be problematic for the model to select the extremely frequent allomorph as the input as found the last two chapters. The assumption that the high-frequency bare form allomorphs are considered to be basic allomorphs, however, first requires PSI-OT-GLA to capture how a variety of stem-final coronal obstruents can be derived from an input stem-final /t/. Such unfaithful mappings are nonetheless impossible in some contexts in PSI-OT-GLA as the markedness constraints that can force a stem-final /t/ to alternate in the suffixed context are dominated by faithfulness constraints. For example, if the input of 'day (acc.)' is /ot+ɨl/, the faithful output [o.t-ɨl] is preferred over the target output [o.s-ɨl] since assibilatory affrication does not occur before /ɨ/ in Korean, as shown in Tableau 5.7. The case might be at best that *Tɨ will be promoted to be on a par with IDENT(cont)-IO in the constraint hierarchy, which can at times give rise to the correct outputs, but a great number of [o.t-ɨl] outputs predicted here is inconsistent with the absence of stem-final /t/ in the production data in Jun & Lee's (2007) experiment (cf. Do 2012; see §4.8 for a discussion).

/ot+ɨl/	IDENT(cont)	*Tɨ	*/+strid/	LAZY
☞☹[o.t-ɨl]		*		**
✓[o.s-ɨl]	*!		*	***

Tableau 5.7. Preference toward the faithful output [o.t-ɨl] with the input /ot+ɨl/; '✓' = correct output, '☹' = output error

The solution lies in the evaluation of whether an unnatural markedness constraint can be created to forbid stem-final /t/ in suffixed nouns by the Tolerance Principle. As reviewed in §2, the dictionary type frequency count in Jun (2010) reveals that none of all 878 suffixed nouns with a stem-final coronal obstruent has a stem-final /t/ (0/878). Based on the Tolerance Principle, the exception number must be lower than 878 / log(878) = 129.5 to generalize the prohibition of stem-final /t/ as an unnatural markedness constraint. Without any exception (i.e. no suffixed noun has a stem-final /t/), the top-ranked unnatural constraint *t/_]$_{STEM}$-S (12) is presumably induced by learners. This top-ranked unnatural markedness constraint always rules out the faithful output of a suffixed noun with an underlying stem-final /t/ as in Tableau 5.8 to be in accord with the variation patterns observed by Jun & Lee.

(12) *t/_]$_{STEM}$-S: A stem-final /t/ is prohibited when preceding a suffix.

/ot+ɨl/	*t/_]$_{STEM}$-S	IDENT(cont)	IDENT(del rel)	*Tɨ
☞[os-ɨl]		*		
☞[otʃ-ɨl]			*	
[ot-ɨl]	*!			*

Tableau 5.8. Stem-final /t/ in suffixed nouns blocked by top-ranked *t/_]$_{STEM}$-S

As stem-final /t/ is absolutely banned in suffixed nouns, the optimal output varies depending on the ranking of faithfulness constraints. Different correct outputs of suffixed noun forms require different mappings between an underlying stem-final /t/ and the target stem-final coronal obstruent. Conflicting re-ranking processes in Tableau 5.9 are thus triggered as the source of stem-final variations. Suppose that the four constraints in Tableau 5.8 have a constraint value of 100, 50, 50, and 0 respectively, IDENT(cont)-IO must be demoted and IDENT(del rel)-IO must be promoted to generate the target output of 'clothes (acc.)' without any exception. By contrast, if the standard pronunciation of 'day (acc.)' [na.tʃ-ɨl] needs to be derived from the input /nat+ɨ[/, IDENT(del rel)-IO must be lower-ranked than IDENT(cont)-IO. If the pressure that requires the dominance of IDENT(del rel)-IO is higher, the input /nat+ɨl/ will surface as [na.s-ɨl] more frequently, and the mapping /t/→[s] is extended from /ot+ɨl/→[o.s-ɨl]. If IDENT(cont)-IO is promoted more often and thus dominates IDENT(del rel)-IO, the input /ot+ɨl/ will surface as [o.tʃ-ɨl]. That is, the mapping /t/→[tʃ] in /ot+ɨl/→[o.tʃ-ɨl] is extended from /nat+ɨl/→[na.tʃ-ɨl].

/ot+ɨl/	*t/_]$_{\text{STEM}}$-S 100	IDENT(cont) 50	IDENT(del rel) 50	*Tɨ 0
☞✓[o.s-ɨl]		*→		
☞☹[o.tʃ-ɨl]			←*	
[ot-ɨl]	*!			*
/nat+ɨl/				
☞☹[na.s-ɨl]			*→	
☞✓[na.tʃ-ɨl]		←*		
[nat-ɨl]	*!			*

Tableau 5.9. Conflicting constraint re-ranking of faithfulness constraints for either a /t/→[s] mapping or a /t/→[tʃ] mapping; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

As shown previously in §2, if a suffixed noun is not produced as its standard pronunciation, it is often produced with a stem-final [s]. The high type frequency of the /t/→[s] mappings, from which a high reliability is converted in the base identification approach, can also help explain how stem-final /s/ becomes prevalent in the current PSI account:[62] The high type frequency (and probably a high total token frequency) implies more attempts to rank IDENT(del rel)-IO higher than IDENT(cont)-IO. The constraint value of IDENT(del rel)-IO may thus be slightly higher than the value of IDENT(cont)-IO. When speakers produce other suffixed nouns by selecting a bare allomorph as the stem input, the surface variants with a stem-final [s] emerge (presumably 70% of the time), as in the case of 'breast (acc.)' and 'flower (acc.)' in Tableau 5.10.

[62] There are 878 noun types with a stem-final coronal obstruent, and 393 of them has a stem-final /s/. The number of exceptions is thus 878-393 = 485, which is much higher than the threshold 878/log(878) = 129.54; a unnatural markedness constraint that forces all stem-final coronal obstruents to be /s/ thus cannot be created.

/tʃət+ɨl/		*t/_]$_{STEM}$-S 100	IDENT(del rel) 53	IDENT(cont) 47
[tʃəs-ɨl]	70%			*
[tʃətʃ-ɨl]	30%		*	
[tʃət-ɨl]	0%	*!		
/k'ot+ɨl/				
[k'o.s-ɨl]	70%			*
[k'o.tʃʰ-ɨl]	30%		*	
[k'o.t-ɨl]	0%	*!		

Tableau 5.10. Extension of /t/→[s] mapping to 'breast (acc.)' and 'flower (acc.)'

Other extensions can also be made possible via a similar re-ranking process. For example, to produce the standard pronunciation of 'field (loc.)/(dat.)' [patʰ-e] with the bare allomorph as the stem input /pat+e/, IDENT(asp)-IO must be demoted to be lower-ranked than IDENT(del rel)-IO and IDENT(cont)-IO as in Tableau 5.11. If the ranking is solidly built, the /t/→[tʰ] mapping can also be extended to other suffixed nouns. The alternation is nevertheless rarely extended due to a lower type frequency (and perhaps a lower total token frequency) reported in Jun (2010); the pressure that moves IDENT(asp)-IO to a higher ranking position is relatively weaker.

/pat+e/	*t/_]$_{STEM}$-S	IDENT(asp)-IO	IDENT(cont)	IDENT(del rel)
☞✓[patʰ-e]		*→		
☞☹[pas-e]			←*	
☞☹[patʃ-e]				←*
[pat-e]	*!			

Tableau 5.11. Constraint re-ranking of faithfulness constraints for a /t/→[tʰ] mapping; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

The possible extensions with different rankings of faithfulness constraints are summarized in Table 5.4, including the estimated frequency of each extension. As mentioned above, the /t/→[s] mapping should have the highest extension frequency, and the other two mappings are less likely to be extended as reported in Jun (2010) and Jun & Lee (2007).

Extension	Ranking	Frequency
/t/→[s]	*t/_]STEM-S » {IDENT(asp), IDENT(del rel)} » IDENT(cont)	High
/t/→[tʃ]	*t/_]STEM-S » {IDENT(asp), IDENT(cont)} » IDENT(del rel)	Low
/t/→[tʰ]	*t/_]STEM-S » {IDENT(del rel), IDENT(cont)} » IDENT(asp)	Low

Table 5.4. Frequency and ranking of possible extensions

Another seemingly possible /t/→[tʃʰ] extension is omitted in the foregoing discussion since /t/→[tʃʰ] is harmonically-bounded by /t/→[tʃ] as in Tableau 5.12.

/t+ɨl/	IDENT(cont)	IDENT(del rel)	IDENT(asp)-IO
☞[tʃ-ɨl]		*	
[tʃʰ-ɨl]		*	*!
[s-ɨl]	*!		

Tableau 5.12. [tʃʰ] harmonically-bounded by [tʃ] with an underlying stem-final /t/

This output gap leaves a problem for the current PSI account since in the experimental results in Jun (2010), in which [tʃʰ-ɨl] was rated as the second highest acceptable stem-final coronal obstruent of accusative forms. Jun & Lee (2007) also show that stem-final [tʃʰ] is frequently adapted as the primary pronunciation of 'field (acc.)', which has a stem-final [tʰ] in its standard pronunciation but surfaces as [patʃʰ-ɨl] eight times. However, it should be noted here that in Jun & Lee's (2007) experimental results, a stem-final [tʃʰ] is possible in a suffixed noun only when (1) the standard pronunciation of the noun already has a stem-final [tʃʰ], and (2) the standard pronunciation has a stem-final [tʰ]. In other words, a surface stem-final [tʃʰ] emerges from either an underlying /tʃʰ/ or /tʰ/. I will return to explain these mappings in §3.3.

3.2 Against stem-final non-plain /p/ and /k/ in suffixed nouns

The variation patterns of stem-final labial and velar stops are more straightforward since they frequently surface as plain stops in suffixed contexts even if they are aspirated or tensed in their standard pronunciation. Following the previous section, bare form allomorphs are assumed to be the 'basic' allomorphs, and the stem-final plain labial and velar stops in bare form allomorphs (e.g. /ip/ 'leaf') surface faithfully in their suffixed form (e.g. [ip-i] 'leaf (top.)'; cf. standard pronunciation [ipʰ-i]). According to the well-formedness survey in Jun (2010), suffixed forms with a stem-final plain labial or velar stop is judged to be the most acceptable forms even if the stem-final stop is aspirated in the standard pronunciation. Similar patterns have been found in Jun & Lee's (2007) elicitation task (see Table 5.3): The stem-final velar stop in the stem /puəkʰ/ 'kitchen' is frequently de-aspirated in all three testing morphological contexts (i.e. [puək-i] 'kitchen (nom.)', [puək-ɨl] 'kitchen (acc.)', and [puək-e] 'kitchen (loc./dat.)'). Albeit being more resistant to de-aspiration, the suffixed noun 'leaf (acc.)' at times surfaces as [ip-ɨl], too.

When bare form allomorphs are selected as the stem input, PSI-OT-GLA can predict the same surface variants since the standard pronunciations with aspirated and tense stops are harmonically-bounded by the faithful outputs as in Tableau 5.13. However, since other surface allomorphs can also be selected as the stem input in PSI-OT-GLA, it is possible for a suffixed noun to surface as its standard pronunciation as in Tableau 5.14, and these outputs should be ruled out for a similar variation pattern to be captured by PSI-OT-GLA.

/puək+ɨl/	IDENT(asp)	IDENT(tense)	*/+asp/	*/+tense/
☞[puə.k-ɨl]				
[puə.kʰ-ɨl]	*!		*	
/pak+ɨl/				
☞[pa.k-ɨl]				
[pa.k'-ɨl]		*!		*

Tableau 5.13. Surface variations generated from bare form allomorphs in the inputs

/puəkʰ+ɨl/	IDENT(asp)	IDENT(tense)	*/+asp/	*/+tense/
[puə.k-ɨl]		*!		
☞[puə.kʰ-ɨl]			*	
/pak'+ɨl/				
[pa.k-ɨl]		*!		
☞[pa.k'-ɨl]			*	

Tableau 5.14. Standard pronunciations generated from other allomorphs in the inputs

The low proportions of standard pronunciations with aspirated and tense stops can be attributed to the extreme type frequencies bias toward suffixed nouns with a stem-final plain labial or velar stop regardless of the suffixal contexts (2,193/2,315 for /p/ and 7,501/7,537 for /k/) as noted in §2,[63] which may thus suggest the induction of top-ranked unnatural markedness constraints in (13) as well.[64]. With a lower-ranked IDENT(asp)-IO and IDENT(tense)-IO, aspirated and tense stops in non-bare allomorphs surface as plain stops at the right stem boundary as illustrated in Tableau 5.15.

(13) Constraints for labial/velar de-aspiration

a. [Lab]/_]$_{\text{STEM}}$=[p]: Stem-final labial stops must be [p] (i.e. plain bilabial stop).

b. [Dor]/_]$_{\text{STEM}}$=[k]: Stem-final labial stops must be [k](i.e. plain velar stop).

/puəkʰ+ɨl/	[Dor]/_]$_{\text{STEM}}$=[k]	IDENT(asp)	IDENT(tense)
☞[puə.k-ɨl]		*	
[puə.kʰ-ɨl]	*!		
/pak'+ɨl/			
☞[pa.k-ɨl]			*
[pa.k'-ɨl]	*!		

Tableau 5.15. De-aspiration of stem-final labial and velar stops

3.3 Locative/dative (and other non-bare) allomorphs as stem inputs

In the foregoing sections, I assumed that bare allomorphs always has a dominant token frequency and thus has a higher chance to be selected as the input; all suffixed nouns must be derived from bare forms. However, locative/dative allomorphs might be strong competitors against bare allomorphs in the input selection process. Unlike the nominative and accusative suffixes, the locative/dative suffix /-e/ is obligatory and can never be omitted even in conversational speech (Jun 2010:171). Therefore, for nouns which are used in the locative/dative context more frequently, their locative/dative allomorph should also have a higher token frequency that allows it to be another preferred input choice. For instance, it is natural to produce the stem 'field' more prevalently in the locative/dative context (see §4.1). Furthermore, the

[63] The sum can be broken into context-specific counting in Jun (2010) as follows: Accusative (-ɨl) = 711/743 for [p] and 2471/2478 for [k]; Topic (-ɨn) = 395/413 for [p] and 1401/1407 for [k]; Directive (-ɨlo) = 370/394 for [p] and 1336/1346 for [k]; Locative/Dative (-e) = 504/534 for [p] and 1562/1575 for [k]; Locative/Source (-esə) = 213/231 for [p] and 731/741 for [k].

[64] The number of exceptions of [Lab]/_]$_{\text{STEM}}$=[p] is 2315-2193 = 122, which is lower than the threshold 2315/log(2315) = 298.8. The number of exceptions of [Dor]/_]$_{\text{STEM}}$=[k] is 7537-7501 = 36, which is also lower than the threshold 7537/log(7537) = 844.2

locative/dative allomorphs are identical to the URs in the standard analysis; they can derive all surface forms. Therefore, while a bare allomorph might be a 'better' input for other stems, the locative/dative allomorph might be in fact recognized as a more 'basic' allomorph for the stem 'field' with its higher token frequency and efficiency of generating more correct outputs.

This essential feature of allowing different allomorphs to be the stem input in PSI can help explain the rare surface variations in the locative/dative context. Recall that in Jun & Lee's (2007) experimental results, locative/dative forms were mostly produced as their standard pronunciation. The form 'day (loc.)/(dat.)', for example, was always produced as [natʃ-e] without the lenition variant [nas-e]. Likewise, the standard pronunciations [k'otʃ^h-e] and [pat^h-e] were also the dominant production of 'flower (loc.)/(dat.)' and 'field (loc.)/(dat.)' respectively.[65] If their locative/dative allomorph is selected as the stem input regularly, the stem-final consonant can naturally be preserved in their locative/dative form (e.g. /pat^h+e/→[pat^h-e] 'field (loc.)/(dat.)').

The assumption that non-bare allomorphs are not always selected as the stem input also helps account for the mystery that stem-final [tʃ^h] variants are possible only when the a stem-final [tʃ^h] or [t^h] occurs in the standard pronunciation of their suffixed forms as noted at the end of §3.1. For example, the stem 'field' has two surface allomorphs /patʃ^h/ and /pat^h/ from the nominative form [patʃ^h-i] and other suffixed forms (e.g. [pat^h-e] 'field (loc.)/(dat.)'). If the former is selected as the stem input in the production of the accusative form (i.e. /patʃ^h+ɨl/), the surface variants with a stem-final [tʃ^h] emerges (i.e. /patʃ^h+ɨl/→[pa.tʃ^h-ɨl]). If the locative/dative allomorph /pat^h/ is selected as the stem input in the same production (i.e. /pat^h+ɨl/) and *Tɨ is somehow promoted higher than IDENT(del rel)-IO (see §3.4 below), assibilatory affrication may occur to generate the same surface variant [pa.tʃ^h-ɨl] as well as in Tableau 5.16.

[65] Similar patterns have been found in Choi's (2004) survey, in which the proportion is higher than 70% for locative/dative forms with an etymologically stem-final /t^h/ to be produced as their standard pronunciation […t^h-e].

/patʰ+ɨl/	*Tɨ	IDENT(del rel)
☞[pa.tʃʰ-ɨl]		*
[pa.tʰ-ɨl]	*!	

Tableau 5.16. Surface variant [pa.tʃʰ-ɨl] from the input /patʰ+ɨl/ with higher-ranked *Tɨ

In sum, the selection of a basic allomorph as the stem input is expected to diverge for each stem with a different contextual distribution as we have seen in the Dutch morphophonological acquisition and the tonal change in Standard Mandarin in previous chapters.

3.4 After the emergence of stem-final variations

This section explores what the gradual grammar and lexical changes will occur in PSI-OT-GLA after stem-final variations are initialized and eventually become the learning inputs of the next generation as laid out in Ch. 4. The basic assumption follows the previous predictions: The /t/→[s] mappings have the highest token and type frequency and thus promote IDENT(del rel)-IO and IDENT(asp)-IO over IDENT(cont)-IO; a stem-final [s] can thus be derived when a bare form allomorph is selected as the stem input of a suffixed noun. However, this ranking also generates the stem-final [s] variations for other suffixed nouns when bare form allomorphs are also selected as the stem input. For example, the input /pat+ɨl/ for 'field (acc.)' may have a higher chance to surface as the variant [pas-ɨl] (Tableau 5.17). Another variant [pa.tʃ-ɨl], along with the standard pronunciation [pa.tʰ-ɨl], is also possible but less frequent.

/pat+ɨl/	*t/_]$_{\text{STEM}}$-S 100	IDENT(asp) 53	IDENT(del rel) 53	IDENT(cont) 47
[pa.s-ɨl] 70%				*
[pa.tʃ-ɨl] 5%			*	
[pa.t^h-ɨl] 15%		*		
[pa.t-ɨl] 0%	*!			

Tableau 5.17. Surface variants from the input /pat+ɨl/ with lower-ranked IDENT(cont)

The word ‘field (acc.)’ now has three different surface forms, and thus three different target outputs when serving as the learning inputs for the learners of the next generation, and among the three target outputs, [pas-ɨl] has the highest token frequency. Recall that at the beginning of the P stage, *Tɨ is initially ranked at the top of the constraint hierarchy. To acquire a constraint grammar that forbids assibilatory affrication before [ɨ] as in Standard Korean, the target output must be [pa.t^h-ɨl] to trigger the demotion of *Tɨ. After variations of suffixed nouns emerge, however, the primary target output becomes [pas-ɨl] and *Tɨ needs not to be demoted below faithfulness constraints to preserve underlying /t^h-ɨ/ or /t-ɨ/ sequences. Consequently, if the stem input has a stem-final /t^h/ before /ɨ/, it will surface as [tʃ^h] or [s] to avoid violating *Tɨ. Since IDENT(cont)-IO is ranked slightly lower than IDENT(del rel)-IO for the /t/→[s] mappings (see §3.1), the [pa.s-ɨl] variant will be more frequent as shown in Tableau 5.18.

/pat^h+ɨl/	*Tɨ 100	IDENT(del rel) 50	IDENT(cont) 47
[pa.s-ɨl] 70%			*
[pa.tʃ^h-ɨl] 30%		*	
[pa.t^h-ɨl] 0%	*!		

Tableau 5.18. Surface variants from the input /pat^h+ɨl/ with higher-ranked *Tɨ

The production /t^h/ before the suffix /i/ in the nominative context is also affected when IDENT(cont)-IO is lowered-ranked than IDENT(del rel)-IO as in Tableau 5.19. The constraint *Ti is still high-ranked as in Standard Korean since /t^h-i/ or /t-i/ combination is still absent after stem-final variations occur, but a lower-ranked IDENT(cont)-IO makes the standard pronunciation with assibilatory affrication less possible.

/pat^h^+i/		*Ti 100	IDENT(del rel) 50	IDENT(cont) 47
[pa.s-i]	70%			*
[pa.tʃʰ-i]	30%		*	
[pa.tʰ-i]	0%	*!		

Tableau 5.19. Surface variants from the input /patʰ+i/ with higher-ranked *Ti

Stem-final [s] in suffixed nouns has two different sources after the above constraint re-ranking – either from a stem-final /t/ in the bare form allomorph or a stem-final /tʰ/ before /i/ or /ɨ/ in the suffix. Note that these grammar changes will not affect stem-final /tʰ/ before the locative/dative suffix /-e/ since no top-ranked constraint bans the [tʰ-e] sequence. Therefore, if the locative/dative allomorph /patʰ/ is frequently selected as the stem input of the locative/dative form as anticipated in §3.3, the standard pronunciation [patʰ-e] can still be produced constantly; the reason why surface variations rarely occur in the locative/dative context in Jun & Lee's (2007) production data can thus be explained.

3.5 Summary of the stem-final variation development

To sum up, the predicted emergence of complex stem-final coronal variations in Korean can be depicted as six successive stages involving a lexical re-analysis as listed in (14). Stem-final non-coronal variations on the hand can be explained by the invention of unnatural markedness constraints forbidding stem-final aspirated and tense non-coronal obstruents. Mappings in Figure 5.2 summarize all possible suffixed noun variations introduced from §3.1 to §3.4 with different stem inputs, and the primary grammar changes with a lower-ranked IDENT(cont)-IO and higher-ranked *Tɨ in this development is summarized in Figure 5.3 with only relevant sub-rankings. These diachronic developments are simulated with PSI-OT-GLA in §4.

(14) Six stages of the Korean stem-final variation development

a. Bare form allomorphs may be selected as the stem input of suffixed nouns more frequently due to their high token frequency.
b. Following (a), a great number of the /t/→[s] mappings re-rank IDENT(cont)-IO to be lower than other faithfulness constraints to allow the standard pronunciation with a stem-final [s] (e.g. /ot+ɨl/→[o.s-ɨl] 'clothes (acc.)').

c. The re-ranking process generates the stem-final [s] variant for other suffixed nouns (e.g. /pat+ɨl/→[pa.s-ɨl] 'field (acc.)' and /pat^{h}+i/→[pa.s-i] 'field (nom.)').

d. When stem-final [s] variants become the learning inputs and targets, more /t/→[s] mappings are required, and the demotion of IDENT(cont)-IO in (b) is further supported.

e. Following (c) and (d), since the /t^{h}+ɨ/ sequences are gradually replaced with the /s+ɨ/ sequences in the learning inputs, it is unnecessary to demote *Tɨ. A higher-ranked *Tɨ then triggers the overall change from /t^{h}+ɨ/ sequences to [s-ɨ].

f. The full cycle from (b) to (e) is repeated with increasing stem-final [s] variants.

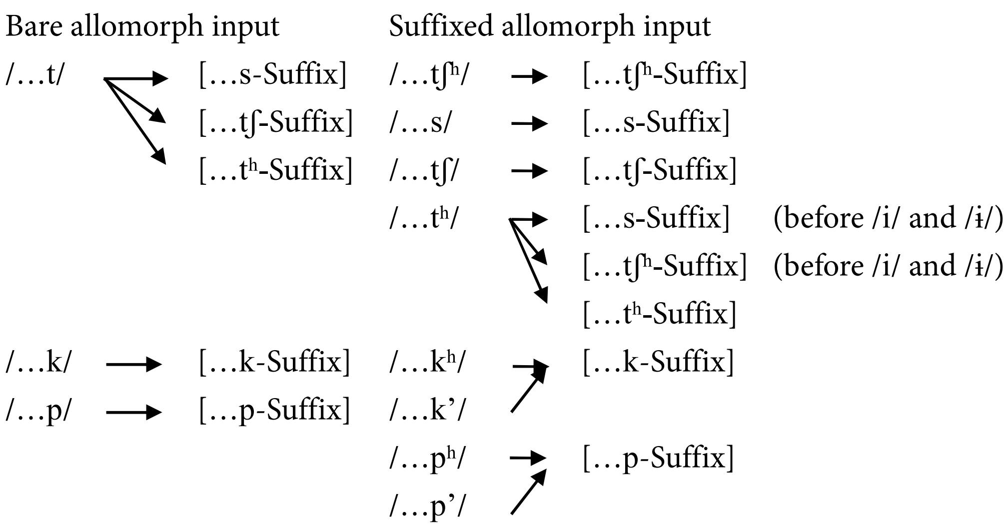


Figure 5.2. Summary of surface variations of suffixed forms from different stem-inputs

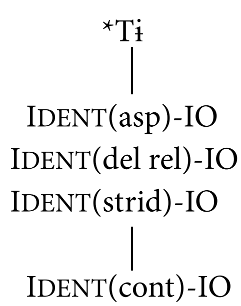


Figure 5.3. Grammar changes with stem-final variations in Korean

4. Modeling Korean stem-final variations with PSI-OT-GLA

The organization of this simulation section starts with the set-up, including the input structure and the distributional probability of individual learning inputs in §4.1, the initial/target constraint grammar and the output candidates for different inputs in §4.2, and the involved lexical factors in §4.3. The perception and production process in every learning cycle are illustrated in §4.4, followed by the design of a process of learning over generations and a simulated elicitation task in §4.5. The overall simulated results generated with the algorithm complied as a Java® program are discussed in §4.6.[66] Potential simulation changes that could bring the simulated results closer to the experimental results are enumerated in §4.7, and the chapter concludes with a comparison between base identification and PSI-OTGLA in §4.8.

4.1 Training corpus, input distribution and input structure

The learning inputs in the following simulation are identical to the thirteen target noun stems in Jun & Lee's (2007) experiment to model the production of each stem in three morphosyntactic contexts (i.e. nominative, accusative, locative/dative). The token frequency of each stem in various morphosyntactic contexts was counted from newspaper and magazine articles and news broadcasting scripts in the Sejong text corpus of 5.5 million words (http://www.sejong.co.kr) in Table 5.5. The four suffixed forms (i.e. topic, nominative, locative/dative, accusative) were chosen in the corpus counting for two reasons. First, surface noun variations were discovered in these contexts in previous studies (e.g. Choi 2004, Kang et al. 2004, Kim 2003), including the three contexts in Jun & Lee's experiment. Second, only the four suffixed forms allow a less arbitrary estimation of their token frequency in speech contexts (see below).

[66] The source code is available at http://hdl.handle.net/10402/era.39161.

Stem	Gloss	bare	Top. (-ɨn)	Nom. (-i)	Loc./Dat. (-e)	Acc. (-ɨl)
nas	‘sickle’	1	2	4	1	10
os	‘clothes’	261	65	135	60	814
natʃ	‘day’	73	1[67]	22	135	3
tʃətʃ	‘breast’	12	1[68]	28	6	118
k’otʃʰ	‘flower’	249	108	289	31	350
natʃʰ	‘face’	6	1	35	1	30
patʰ	‘field’	26	16	23	62	81
pʰatʰ	‘red bean’	6	0	2	0	4
puəkʰ	‘kitchen’	69	16	24	34	29
pak’	‘outside’	68	41	29	360	156
pjək	‘wall’	125	21	49	306	202
ipʰ	‘leaf’	34	11	74	6	69
pap	‘rice’	209	41	38	14	440
Total		1139	324	752	1016	2306

Table 5.5. Token frequency of thirteen Korean noun stems in five morphosyntactic contexts

The distributional probabilities of each learning input should have been directly converted from the raw token frequency as in the previous Dutch and Mandarin simulations. However, the text token frequencies are not necessarily similar to those in child-directed speech (CDS). In the Korean case, estimating the token frequency of morphologically complex forms is particularly difficult since suffixes that appear in texts might be frequently omitted in conversational speech (e.g. topic and nominative markers). Thus, the token frequencies in some morphosyntactic contexts might be overestimated. Lee’s (1999) acquisition study also indicates a significantly higher bare form proportion (75%) than the proportion of nominative (20%) and other (5%) suffixed forms in CDS. Without a corpus documenting how parents deliver learning inputs to Korean-learning children, I followed the plain description in Lee (1999) to invent naïve correction guidelines to correct the raw token frequency of different suffixed nouns. The raw frequency of locative/dative forms, whose suffix is never omitted, was not corrected. With only

[67] The form [na.tʃ-ɨn] is used to mean ‘low’ more frequently, and it is a rare case to topicalize the stem ‘day’.

[68] Like the case of [na.tʃ-ɨn], the form [tʃə.tʃ-ɨn] almost always expresses the meaning ‘wet’ instead of ‘breast (top.)’.

20% of nominative forms in CDS, the chance for dropping the nominative suffix was assumed to be 80%; that is, the raw token frequency decreased by 80%. Likewise, for topic and accusative forms, the raw token frequency decreased by 95%. The corrected token frequencies are summarized in Table 5.6 with the corresponding and the parenthesized corresponding distributional probabilities.

Stem	Gloss	bare	Top. (-ɨn)	Nom. (-i)	Loc./Dat. (-e)	Acc. (-ɨl)
nas	‘sickle’	1 (0.04%)	0.1 (≈0%)	0.8 (0.03%)	1 (0.04%)	0.5 (0.02%)
os	‘clothes’	261(10.71%)	3.25 (0.13%)	27 (1.11%)	60 (2.46%)	40.7 (1.67%)
natʃ	‘day’	73 (3%)	0.05 (≈0%)	22 (0.18%)	135 (5.54%)	0.15 (0.01%)
tʃətʃ	‘breast’	12 (0.49%)	1 (≈0%)	5.6 (0.23%)	6 (0.25%)	5.9 (0.24%)
k’otʃ^h	‘flower’	249 (10.22%)	5.4 (0.22%)	57.8 (2.37%)	31 (1.27%)	17.5 (0.72%)
natʃ^h	‘face’	6 (0.25%)	0.05 (≈0%)	7 (0.29%)	1 (0.04%)	1.5 (0.06%)
pat^h	‘field’	26 (1.07%)	0.8 (0.03%)	4.6 (0.19%)	62 (2.54%)	4.05 (0.17%)
p^{h}at^h	‘red bean’	6 (0.25%)	0 (0%)	0.4 (0.02%)	0 (0%)	0.2 (0.01%)
puək^h	‘kitchen’	69 (2.83%)	0.8 (0.03%)	4.8 (0.2%)	34 (1.4%)	1.45 (0.06%)
pak’	‘outside’	68(2.79%)	2.05 (0.08%)	5.8 (0.24%)	360 (14.77%)	7.8 (0.32%)
pjək	‘wall’	125 (5.13%)	1.05 (0.04%)	9.8 (0.4%)	306 (12.56%)	10.1 (0.41%)
ip^h	‘leaf’	34 (1.4%)	0.55 (0.02%)	14.8 (0.61%)	6 (0.25%)	3.45 (0.14%)
pap	‘rice’	209 (8.58%)	2.05 (0.08%)	7.6 (0.31%)	14 (0.58%)	22 (0.9%)
Total		1139 (46.74%)	16.2 (0.67%)	150.4 (6.17%)	1016 (41.69%)	115.3 (4.73%)

Table 5.6. Corrected token frequency and distributional probability (parenthesized) of thirteen Korean noun stems in five morphosyntactic contexts

After correction, the proportion of ‘other’ suffixed forms (i.e. topic and accusative) is 0.007% + 4.731% = 4.738%, which is only slight lower than the reported 5% in Lee’s (1999) study. The proportion of nominative forms is an underestimated 6.172%, which may be considered variation specific to this subset of learning inputs. The major difference lies in the lower proportion of bare forms (46.74%). This is because the token frequency of locative/dative forms does not have to be corrected, and the great token number of ‘day (loc./dat.)’, ‘outside (loc./dat.)’, and ‘wall (loc./dat.)’ thus lowers the proportion of bare forms. If the three stems are excluded, the proportion of bare forms increases to (65.7%). Nevertheless, as I proposed in §3, the surface variations patterns might also be attributed to individual distributional variations. The three stems were therefore included in the simulation rather than excluded merely in order to approach the general distributional tendency reported in Lee's study. Furthermore, for the majority of the stems, the token number of bare forms still exceeds that of locative/dative forms, following the general tendency that bare forms are commonly used.

The phonetic form of each stem was also modified to capture the generalization that obstruent contrasts are preserved in the onset position. Among the thirteen stems, the onset aspiration contrast only appears in the allomorph /pʰatʰ/ of the stem 'red bean', the onset tense contrast only appears in the allomorph /k'otʃʰ/ of the stem 'flower', and the onset affricate contrast only appears in the allomorph /tʃətʃ/ of the stem 'breast'. Neither the fricative onset /s/ nor the aspirated affricate onset /tʃʰ/ is present in the learning inputs, with which learners might end up acquiring a constraint grammar that does not preserve these onset contrasts. Since the full contrasts appear in the stem-final position in the selected stem, the phonetic form of the standard pronunciation of each stem was changed by replacing the onset with the stem-final consonant (e.g. [nas]=[sas]) or copying the stem-final consonant as syllable onset (e.g. [os]=[sos]) to avoid the above grammar learning issue.[69] The vowel differences were also ignored by replacing all the stem vowels with V to simplify the input types and thus the process of constructing possible output candidates. The labial inputs can also be merged with the velar inputs since the alternation is essentially identical for the two places of articulation. In the end, phonetic forms representing each stem are listed in (15).

(15) Alternative phonetic forms for the stems as the learning inputs

/sVs/	=	/nas/	'sickle'	/os/	'clothes'
/tʃVtʃ/	=	/natʃ/	'day'	/tʃətʃ/	'breast'
/tʃʰVtʃʰ/	=	/natʃʰ/	'face'	/k'otʃʰ/	'flower'
/tʰVtʰ/	=	/patʰ/	'field'	/pʰatʰ/	'red bean'
/kʰVkʰ/	=	/puəkʰ/	'kitchen'	/ipʰ/	'leaf'
/k'Vk'/	=	/pak'/	'outside'		
/kVk/	=	/pjək/	'wall'	/pap/	'rice'

[69] With this input structure, the token frequency of individual onsets is always identical to the token frequency of the corresponding codas, which might be less unlikely in real learning inputs. The following simulation thus will not seek to capture and explain the acquisition order of Korean onset contrasts at the stage of phonotactic learning.

Note that although different stems may be represented by the same phonetic form, the chance for the stems to be produced by the 'adult' speaker or learners is still different due to different distributional frequencies (see a similar simulation design in Ch. 3). The topic and accusative markers were also simplified as /-ɨ/ for the same suffix-initial vowel.

4.2 Constraint set and initial state

There were seventeen constraints involved in the simulation, including nine markedness constraints in (16), six faithfulness constraints in (17), and three unnatural markedness constraints in (18). The markedness constraint *Tu is excluded since the training corpus does not have any entry with a suffix beginning with /-u/. The same initial Markedness » IO-Faithfulness ranking bias applied to place markedness constraints at the top of the initial ranking as in (19), and the target constraint ranking for standard pronunciations is repeated in Figure 5.4.

(16) Nine markedness constraints in the simulation

*+asp/_]σ, *+strid/_]σ, *+tense/_]σ, *+asp, *+strid, *+tense, *Ti, *Tɨ, LAZY

(17) Five faithfulness constraints in the simulation

IDENT(del rel)-IO, IDENT(cont)-IO, IDENT(asp)-IO, IDENT(tense)-IO, IDENT(strid)-IO, IDENT(σ_1)-IO

(18) Two unnatural markedness constraints in the simulation

*t/_]$_{STEM}$-S, [Dor]/_]$_{STEM}$=[k]

(19) Initial constraint ranking

{*+asp/_]σ, *+strid/_]σ, *+tense/_]σ, *+asp, *+strid, *+tense, *Ti, *Tɨ, LAZY} » {IDENT(del rel), IDENT(cont), IDENT(asp), IDENT(tense), IDENT(strid), IDENT(σ_1)}

{*+strid/_]σ *+asp/_]σ *+tense/_]σ IDENT(σ_1)-IO}

|

*Ti

|

{IDENT(tense)-IO IDENT(asp)-IO IDENT(strid)-IO IDENT(cont)-IO}

|

IDENT(del rel)-IO

|

{*+asp *+strid *+tense LAZY} *Tɨ

Figure 5.4. Target constraint ranking for standard pronunciations in Korean

The unnatural markedness constraints, as proposed in §3.1 and §3.2, build upon the type frequency count from learning inputs. Nevertheless, the distribution of stem-final obstruent types in the training corpus does not reflect the general trend that [s] and [k] are highly dominant stem-final coronal and dorsal obstruents of Korean nouns (see §4.1), preventing unnatural constraints from being projected along the simulation course. The unnatural markedness constraints are therefore assumed to exist by default in the following simulation since the distributional tendency and the corresponding unnatural constraints should emerge in a larger training corpus collecting more noun forms.

Since the unnatural markedness constraints refer to morphological structure, they are not active in the P stage simulation. At the beginning of the M stage simulation, they are assigned a constraint value of 100 to be top-ranked in the existing constraint hierarchy. Note that the unnatural markedness constraint [Lab]/_]$_{STEM}$=[k] was not included due to the simplified input structure which merged labial and velar stops as velar stops in §4.1. This set of constraints will then generate a group of non-harmonically-bounded output candidates from different input types.

Constraint promotion and demotion were also implemented via the same demotion bias (see §1.2 of Ch. 2) on a base plasticity of 0.1. As unnatural markedness constraints were directly created by lexical generalizations, it was assumed that their constraint value was not affected by the promotion and demotion processes. The same evaluation noise randomly generated from a Gaussian distribution with a standard deviation of 2 was also added to the raw constraint value during the probabilistic ranking process.

4.3 Lexical factors

As in the Dutch simulation in Ch. 3, the lexical factors token frequency, Individual Error Proportion (IEP), and Morphological Error Proportion (MEP) were included in the simulation. In particular, an MEP could be induced from each morphosyntactic context as in (20) to give rise to a morphological privilege for a set of surface allomorphs from the same context.

(20) Morphological Error Proportion for each of the four morphosyntactic contexts
MEP[X+Ø]$_{\text{BARE}}$, MEP[X+ɨn]$_{\text{TOP}}$, MEP[X+i]$_{\text{NOM}}$, MEP[X+ɨl]$_{\text{ACC}}$, MEP[X+e]$_{\text{LOC/DAT}}$

The memory decay rate was set to a decrease of 0.0005 to each token frequency per learning cycle, which aimed at lowering the lexical stability of the forms with a distributional probability lower than 0.05% (or a token frequency lower than 1 after correction).

4.4 Perception and production in two-stage learning

As in previous simulations, each learning cycle beings with the perception of learning inputs. That is, the ‘adult’ speaker will randomly select a noun based on the input distribution in Table 5.5 in §4.1, and learners perceives this input and calculate the noun distribution in their lexicon. The adult speaker of Generation 0 is assumed to always produce the standard pronunciation of each noun forms to simulate the state before the diachronic development as in Ch. 4. The learners of Generation 1 will thus always perceive the same input for each noun form. After stem-final variations occur, the learners of following generations may perceive different inputs for the same noun form.

In the P stage simulation, morphological decomposition does not occur during perception, which means that the token frequency of surface allomorphs is not recorded, and lexical factors are not calculated either. In the M stage simulation, the learners start tracking the token frequency of surface allomorphs in different contexts in the learning inputs and calculating IEPs, MEPs, and SPs. The tracking process will create a lexical factor matrix like Table 5.7 for each noun stem with all possible surface allomorphs and the five morphosyntactic contexts.

Stem: ‘field’	MEP[X+Ø]$_{BARE}$	MEP[X+ɨn]$_{TOP}$	MEP[X+i]$_{NOM}$	MEP[X+ɨl]$_{ACC}$	MEP[X+e]$_{LOC/DAT}$
pat^h	0/0	40/0.05	0/0	20/0.05	100/0.05
$patʃ^h$	0/0	0/0	45/0.1	0/0	0/0
pat	350/0.5	0/0	0/0	0/0	0/0

Table 5.7. An example of lexical factor matrix of the stem ‘field’ and the possible surface allomorphs in its standard pronunciations; each cell in the grid represents raw token number/IEP of the surface allomorphs in different contexts

Since different allomorphs may appear in the same context, and the same allomorph may appear in different contexts, each surface allomorph in each context has its own token frequency and IEP. For example, the allomorph /pat/ only occurs in the bare form of ‘field’, and thus only has a high token frequency and a high IEP (due to a higher proportion of output errors derived from a bare form allomorph) in this bare form context. On the other hand, the allomorph /pat^h/ can occur in the topic, accusative and locative/dative forms. It thus has three different low token frequencies and IEPs across the three contexts due to the less frequent use of suffixed forms. The low IEPs can be attributed to the fact that the allomorph almost always surface as the standard pronunciation when being selected as the stem input of every noun form.

The SP of each surface allomorph in the same column (i.e. same morphosyntactic context) is then calculated with the corresponding MEP. In the case of the allomorph /pat^h/, $SP_{/pat^h/\text{-}TOP}$, $SP_{/pat^h/\text{-}ACC}$, and $SP_{/pat^h/\text{-}LOC/DAT}$ are calculated with MEP[X+ɨn]$_{TOP}$, MEP[X+ɨl]$_{ACC}$, and MEP[X+e]$_{LOC/DAT}$, respectively, and the sum of the three individual SPs is equal to $SP_{/pat^h/}$ of the stem ‘field’.

In the production turn, the learners first randomly select a noun form to produce based on the noun distribution in the lexicon. In the P stage simulation, an input is always identical to the perceived adult forms. If there is more than one form for the selected noun when stem-final variations occur, the chance for selecting one form as the input of the noun is identical to the chance it is perceived. In the M stage simulation, the stem input is probably selected from a set of surface allomorphs by referring to the stem’s lexical factor matrix as shown above.

Finally, if a noun form is always produced as the same phonetic form in the learning inputs, it will be the only possible target output. If stem-final variations are produced in the learning inputs for the same noun form, the chance for each variant to be the target output equals to the probability for the variant to occur in the learning inputs.

4.5 Learning over generations and a simulated elicitation task

The simulation process of a diachronic development is identical to the design of the Mandarin simulation in Ch. 4. First, the Generation 0 speaker is assumed to always produce the standard pronunciation of each form as assumed above, and the Generation 1 learner acquires a constraint grammar and adjusts lexical factors based on such learning inputs. By the end of the Generation 1 morphophonological acquisition, the learner becomes the 'adult' speaker and produces learning inputs with the final constraint grammar and lexical factors for the Generation 2 learner.

In each learning generation, the learning cycle repeated 10,000 times in the both P and M stage simulation. Due to a highly scattered learning input distribution, the number of P stage learning cycles has been increased by ten times to ensure that each noun form can be sufficiently perceived by the learners during the P stage simulation to learn native phonotactics correctly from the very beginning of the simulation. By the end of the M stage simulation, a simulated elicitation task was conducted to produce the thirteen stems 100 times in the nominative, accusative, and locative/dative contexts with the constraint values and lexical factors at the end of the morphophonological acquisition. Each learning generation was repeated ten times as in Ch. 4, and the results, including constraint values, lexical factors, and productions in the elicitation task were averaged across the ten repetitions for a typical learning pattern of a learning generation. The number of learning generations was also set to ten to allow an observation of a long-term diachronic development.

4.6 Results

The following discussion will focus on the differences between the learning results of Generation 1 (i.e. the onset of the diachronic development) and the results of the learning generation which best approximate the experimental results in Jun & Lee (2007). Since stem-final [s] was predicted as the most prevalent variation for suffixed noun forms during the diachronic development, the proportion of stem-final [s] variants should be a valid indicator of the similarity between the simulation and experimental results. Excluding the seven stems ('sickle', 'clothes', 'kitchen', 'leaf', 'outside', 'wall', and 'rice') whose stem-final consonant is already [s] in the standard pronunciation of their suffixed forms or is not a coronal obstruent, the proportion of stem-final [s] variants in Jun & Lee's experimental results is 56 / 173 = 0.324. Based on the chronological change in the proportion in Figure 5.5, Generation 5 has a most similar proportion of 0.303. Therefore, the development through Generation 5 and whether the variation patterns produced in Generation 8 are similar to those in Jun & Lee's experiment will be two primary topics below.

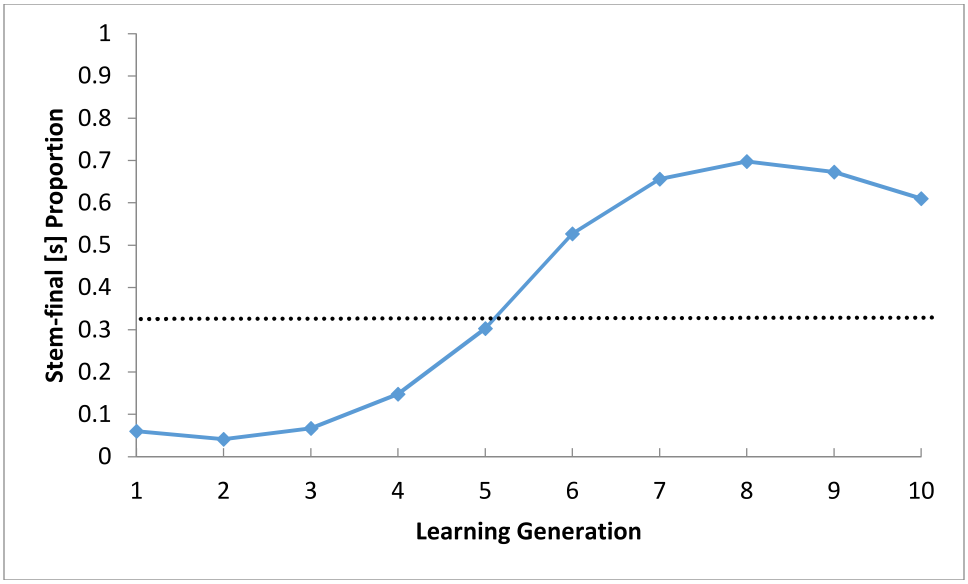


Figure 5.5. Stem-final [s] proportions over ten learning generations; the proportion in Jun & Lee's experiment (0.324) is represented by a dotted line

4.6.1 Grammar development for stem-final coronal variations

The grammar development is examined ahead of other simulated results to investigate whether the constraint grammar for standard pronunciations is correctly acquired at the starting point of the simulation and whether the constraint grammar changes in accord with the predictions made in §3.

Figure 5.6 illustrates the constraint values and the corresponding constraint ranking at the end of the M stage simulation in Generation 1. For stem-final coronal variations, only the constraints *Tɨ, IDENT(strid)-IO, IDENT(cont)-IO, IDENT(del rel)-IO, and IDENT(asp)-IO, which are involved in the predicted grammar change, are discussed here to simplify the discussion. Following the same criterion in Ch. 3 and 4, one constraint is considered to strictly dominate another constraint if the difference in their constraint value is equal or higher than five. The constraint ranking is in general very similar to the target ranking for standard pronunciation as *Tɨ is nearly fully dominated by IDENT(strid)-IO (represented with a dotted line). Thus, the mappings /t/→[s], /t/→[tʃ], /tʰ/→[s], and /tʰ/→[tʃʰ] are generally not possible before /ɨ/ at this stage. What slightly differs from the target ranking is the distance between IDENT(cont)-IO and IDENT(del rel)-IO: While the constraint value of the former is still higher than the value of the latter to derive assibilatory affrication before /i/, the difference between the two values is not significant enough to generate /t/→[tʃ] and /t^h/→[$tʃ^h$] mappings before /i/ without any exception. This is because the higher number of the /t/→[s] mapping when selecting bare form allomorphs as the stem input already influences grammar learning at this stage. The higher-ranked IDENT(asp)-IO bans the /t/→[t^h] mapping in general, and its relatively high constraint value will be discussed in §4.6.2.

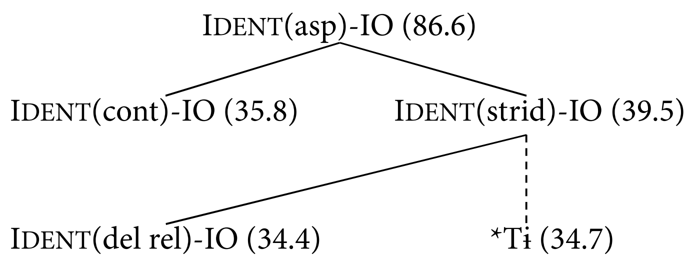


Figure 5.6. Constraint values of five constraints and the corresponding constraint ranking at the end of the M stage simulation in Generation 1

By the end of Generation 5, the grammar changes occur as demonstrated in Figure 5.7. One noticeable difference is that *Tɨ is no longer dominated by faithfulness constraints. Rather, it has a slightly higher constraint value than IDENT(strid)-IO to trigger the mappings /t/→[s], /t/→[tʃ], /t^h/→[s], and /t^h/→[$tʃ^h$] at times. As predicted in §3.4, selecting bare form allomorphs as the stem input may result in the mappings /t/→[s] and /t/→[tʃ] before all noun suffixes, including the topic and accusative markers /-ɨn/ and /-ɨl/ (e.g. /pat+ɨl/→[pa.s-ɨl] 'field (acc.)'). These surface variants will eventually become learning inputs and the pressure to demote *Tɨ below faithfulness constraints reduces. Another significant grammar change is that the distance between IDENT(cont)-IO and IDENT(del rel)-IO further shrinks from 1.4 in Generation 1 to 1 in Generation 5. The change can be attributed to the gradually increasing number of the /t/→[s] mappings, which forms a greater pressure to rank IDENT(del rel)-IO higher than IDENT(cont)-IO.

IDENT(asp)-IO (7.6)
IDENT(strid)-IO (5.5)   *Tɨ (7)

IDENT(cont)-IO (1.8)
IDENT(del rel)-IO (0.8)

Figure 5.7. Constraint values of five constraints and the corresponding constraint ranking at the end of the M stage simulation in Generation 5

The development of the five constraint values over the ten generations is illustrated in Figure 5.8. In the grammar development, it should be obvious that *Tɨ eventually becomes top-ranked after Generation 5. As explained above, the [tʰ-ɨ] sequence of topic and accusative forms in the learning inputs is gradually replaced with the [s-ɨ] and [tʃ-ɨ] sequences after the emergence of the stem-final variation. For learners before Generation 6, the number of [tʰ-ɨ] sequences, albeit decreasing, in the learning inputs is still sufficient for them to demote *Tɨ to be at least unranked with faithfulness constraints. For learners of Generation 6, however, the rarity of the [tʰ-ɨ] sequence in the learning inputs no longer build enough pressure to significantly demote *Tɨ. When the Generation 6 learners become 'adults' with a higher-ranked *Tɨ, [tʰ-ɨ] sequences are totally absent in their production (i.e. the learning inputs for Generation 7). In the end, the demotion of *Tɨ is completely unnecessary, causing the seemingly abrupt change in the constraint ranking after Generation 6.

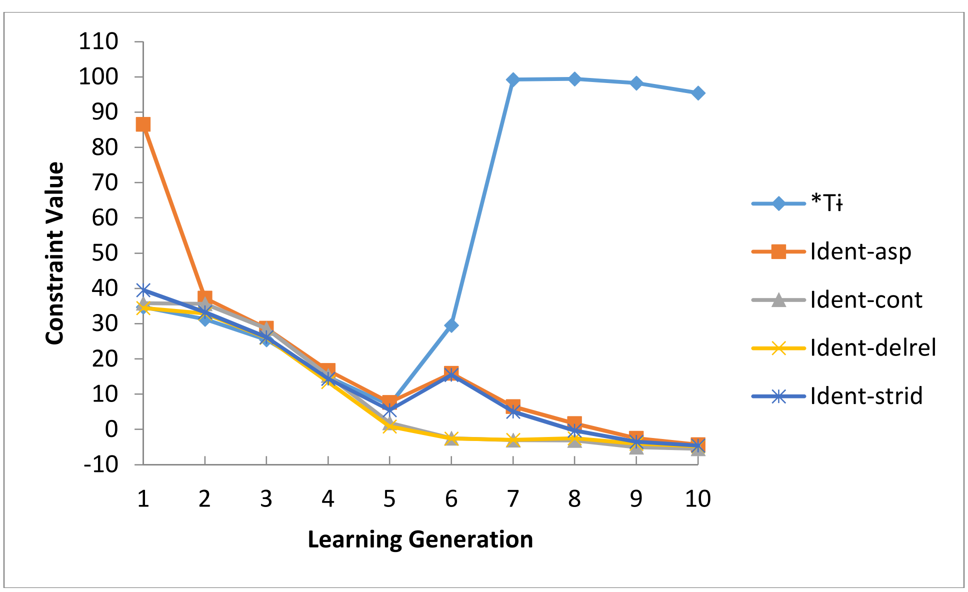


Figure 5.8. Development of constraint values of the selected five constraints

4.6.2 Grammar development for stem-final labial/velar variations

In §3.3, it was assumed by default that stem-final labial/velar stops are produced consistently as plain stops since *t/_]$_{\text{STEM}}$-S and [Dor]/_]$_{\text{STEM}}$=[k] are top-ranked and not involved in constraint promotion/demotion. The fixed position of the unnatural markedness constraints nonetheless does not prohibit the promotion of other faithfulness constraints for the correct outputs. For example, when non-bare form allomorphs are selected as the stem input like /kʰVkʰ+i/, the top-ranked [Dor]/_]$_{\text{STEM}}$=[k] allows the de-aspirated variant *[kʰV.k-i] to be the optimal output as in Tableau 5.20. Since the target output is still the standard pronunciation [kʰV.kʰ-i], a promotion-only constraint re-ranking process is triggered. That is, although [Dor]/_]$_{\text{STEM}}$=[k] remains in the same position as hypothesized, IDENT(asp)-IO will be gradually promoted, and a similar promotion process also applies to IDENT(tense)-IO to derive the standard pronunciation like [k’V.k’-i] from /k’Vk’+i/. Consequently, IDENT(asp)-IO and IDENT(tense)-IO are promoted more frequently than other faithfulness constraints in Generation 1, and IDENT(tense)-IO is even promoted to strictly dominate [Dor]/_]$_{\text{STEM}}$=[k] as shown in Figure 5.9.

/kʰVkʰ+i/	[Dor]/_]$_{\text{STEM}}$=[k]	IDENT(asp)-IO
☞☹[kʰV.k-i]		←*
✓[kʰV.kʰ-i]	*!	

Tableau 5.20. Promotion of IDENT(asp)-IO against [Dor]/_]$_{\text{STEM}}$=[k]; ‘✓’ = correct output, ‘☹’ = output error, ‘←’ = promotion

IDENT(tense)-IO (106.2)
|
[Dor]/_]$_{\text{STEM}}$=[k] (100)
*t/_]$_{\text{STEM}}$-S (100)
|
IDENT(asp)-IO (86.6)

Figure 5.9. Constraint values of four constraints and the corresponding constraint ranking at the end of the M stage simulation in Generation 1

Recall that bare form allomorphs generally have a higher token frequency and selecting bare form allomorphs with a stem-final plain labial/velar stop as the stem input can only surface faithfully (e.g. /kʰVk+i/→[kʰV.k-i]). These stem-final [p]/[k]

variants also became target outputs, which weaken the pressure of promoting IDENT(tense)-IO and IDENT(asp)-IO. IDENT(asp)-IO thus significantly drops to 30 and IDENT(tense)-IO to 27, as shown in the constraint ranking at the end of Generation 5 in Figure 5.10.

[Dor]/_]$_{STEM}$=[k] (100)
*t/_]$_{STEM}$-S (100)
|
IDENT(tense)-IO (74.5)
IDENT(asp)-IO (7.1)

Figure 5.10. Constraint values of four constraints and the corresponding constraint ranking at the end of the M stage in Generation 5

4.6.3 Development of Morphological Error Proportion

The development of MEP in Figure 5.11 is first discussed with the comparison between MEP[X+Ø]$_{BARE}$ and the mean of other MEPs to demonstrate the gradual shift of the 'basic' allomorph from non-bare form allomorphs to bare-form allomorphs. As in the standard analysis, only non-bare form allomorphs can derive all standard pronunciations of Korean nouns, and a significantly higher MEP[X+Ø]$_{BARE}$ in Generation 1 verifies the difference. Thus, as non-bare form allomorphs are selected as the stem input more frequently, a grammar similar to the target grammar can thus be acquired in Generation 1 as revealed in §4.6.1 and §4.6.2.

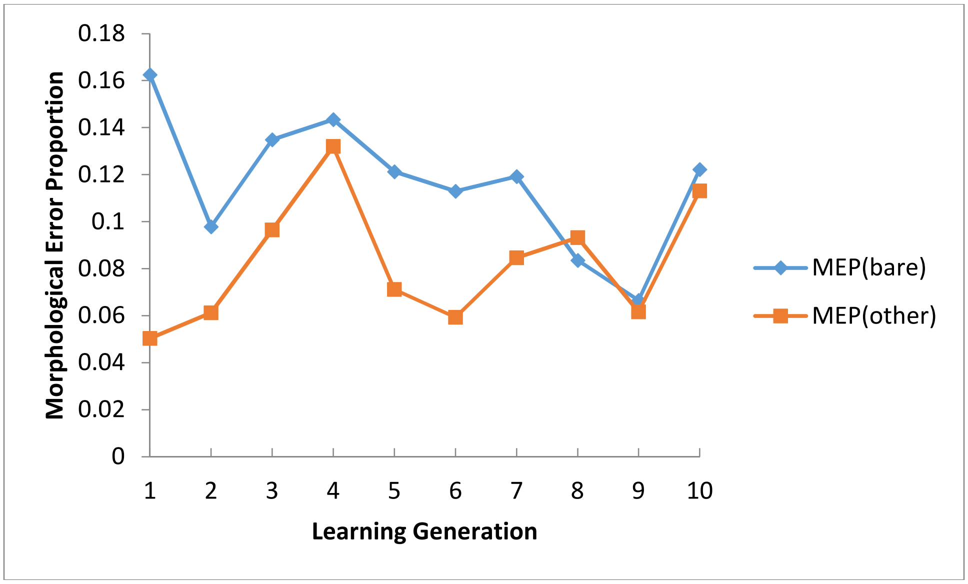


Figure 5.11. MEP[X+Ø]$_{\text{BARE}}$ and the mean of other MEPs over ten learning generations

The emergence of surface variants in the following generations, primarily the ones with a stem-final [s], causes non-bare form allomorphs to generate fewer correct outputs than in Generation 1. For example, assuming that there are two surface accusative variants [pa.tʰ-ɨl] and [pa.s-ɨl] in the learning inputs, selecting a non-bare form allomorph as the stem input such as /pa.tʰ-ɨl/ can only generate one of the two targets depending on the constraint ranking. If the optimal output is [pa.tʰ-ɨl] but the target output is [pa.s-ɨl], an output error occurs. The MEP of non-bare form allomorphs thus climbs up to be closer to MEP[X+Ø]$_{\text{BARE}}$ between Generation 1 and 4.

After Generation 5, however, the MEP of non-bare form allomorphs drops to be significantly lower than MEP[X+Ø]$_{\text{BARE}}$ once again, which can be attributed to an increased number of stem-final [s] variants of suffix forms. When standard pronunciations (e.g. [pa.tʰ-ɨl]) are largely replaced with stem-final [s] variants (e.g. [pa.s-ɨl]), stem-final [s] allomorphs have a higher token frequency in the learning inputs, which are thus selected by learners of the next generation as stem inputs more frequently. Furthermore, the stem-final [s] variants also become the primary target outputs for the learners, which can be derived from the stem-final [s] allomorphs (e.g. /pa.s-ɨl/→[pa.s-ɨl]). Selecting non-bare form allomorphs as stem-input therefore generates fewer output errors as reflected in the lowered MEP. Although the MEP of non-bare form allomorphs never noticeably exceeds

$MEP[X+Ø]_{BARE}$, a smaller difference toward the last generation gradually prevents learners from countering against the general token frequency bias toward bare form allomorphs. The 'basic' allomorph thus gradually shifts to bare form allomorphs for some noun stems as suggested in the following sections.

4.6.4 Similarities between simulated and experimental results

The discussion of the simulated results first concentrates on the stems whose output patterns are similar to those in the experimental results. The simulated results are considered to be similar to the experimental results if the primary and secondary outputs are essentially identical in the both results and if the proportions of these outputs are similar as well (see Appendix D for the detailed output distributions in the both results). The SP will be discussed together with the output patterns, but IEP will be ignored for spatial convenience.

4.6.4.1 Stem: 'sickle'

The stem with similar output patterns to be discussed here is 'sickle' which has a stem-final /s/ in its suffixed forms such as [na.s-i] 'sickle (nom.)'. The simulated results in Generation 5 are listed along with Jun & Lee's experimental results in Table 5.8. The standard pronunciation of the stem's suffixed forms was dominant in Jun & Lee's experiment, and this majority of the standard pronunciation is also captured in the production patterns in Generation 5. Sporadic stem-final [tʃ] outputs are derived from a stem-final [t] when the stem's bare-form allomorph is selected as the stem input. Note that by the end of Generation 5, the constraint value of IDENT(del rel)-IO is not significantly higher than the value of IDENT(cont)-IO; the mappings of /t/→[tʃ] can thus occur as the secondary output in the simulation.

'sickle'	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[nas-i] 100%		[nas-ɨl] 100%		[nas-e] 100%	
Sim	[nas-i] **97%**	[natʃ-i] **3%**	[nas-ɨl] **95.9%**	[natʃ-ɨl] **2.7%**	[nas-e] **96.7%**	[natʃ-e] **2.3%**

Table 5.8. Percentages of the primary and secondary outputs of 'sickle' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs

The SP development of the allomorphs of 'sickle' is visualized in Figure 5.12,[70] which suggests that the allomorph [nas] is still the primary input option along the simulation course. After frequency correction, the stem 'sickle' has one single bare form allomorph and 0.1 + 0.8 + 1 + 0.5 = 2.4 non-bare form allomorphs, which is a ratio of 1:2.4. That is, the frequency bias is actually toward non-bare form allomorphs, which can correctly surface as the standard pronunciation across different contexts. Therefore, the frequency bias and a lower number of output errors (thus a low IEP) naturally lead to a higher SP of the allomorph [nas]. In Generation 5, the SP of [nas] is around 0.94, which means that roughly 94% of the standard pronunciation [na.s-Suffix] are derived faithfully from the input /nas+Suffix/, and the other 2~3% of the standard pronunciation are derived from the input /nat+Suffix/.

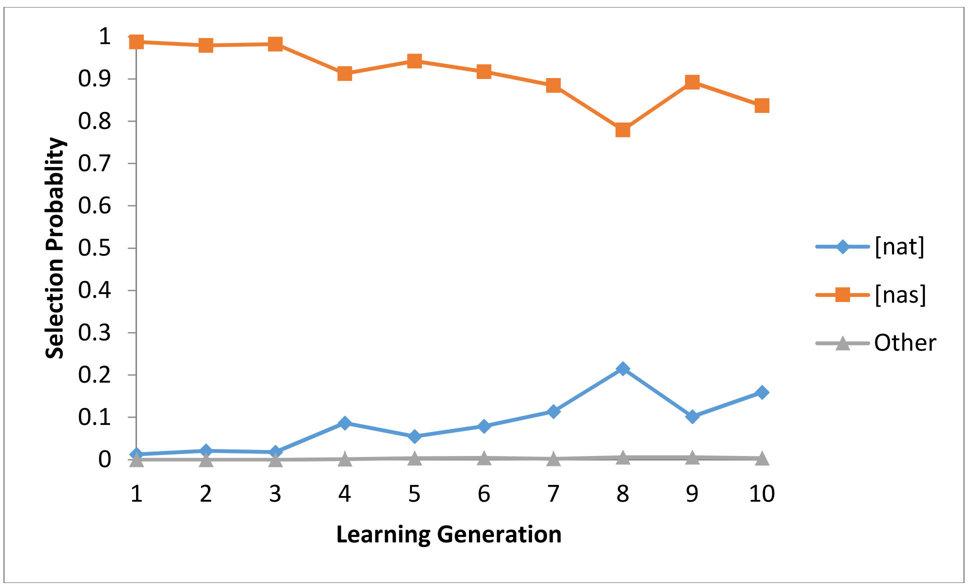


Figure 5.12. SP development of surface allomorphs of 'sickle' over ten learning generations

4.6.4.2 Stem: 'day'

For the stem 'day', PSI-OT-GLA in general correctly predicts that the standard pronunciation with a stem-final [tʃ] is the most preferred production in all three contexts and that if surface variation occurs, the stem-final [s] variant is always the secondary output as in Table 5.9.

[70] Recall that the SP of an allomorph is the sum of the SP of the allomorph in each of the five contexts (see §4.4).

'day'	nominative -*i*		accusative -*ɨl*		locative/dative -*e*	
Exp	[natʃ-i] 90%	[nas-i] 10%	[natʃ-ɨl] 60%	[nas-ɨl] 40%	[natʃ-e] 100%	
Sim	[natʃ-i] **61.4%**	[nas-i] **36.2%**	[natʃ-ɨl] **56.9%**	[nas-ɨl] **34.8%**	[natʃ-e] **56.8%**	[nas-e] **34.3%**

Table 5.9. Percentages of the primary and secondary outputs of 'day' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs.

With a higher number of standard pronunciations, the simulated results suggest that non-bare form allomorphs [natʃ] are still selected as the stem input more frequently. As shown in Figure 5.13, $SP_{/natʃ/}$ is still most dominant in Generation 5, which allows about a chance of 61% for /natʃ/ to be selected as the stem input. However, with the emergence and a gradually increasing number of stem-final [s] variants (e.g. [na.s-i]), the chance for the alternative non-bare form allomorph /nas/ to be selected as the stem input soars. The bare allomorph /nat/ of the stem is never the preferred input choice with a frequency bias toward the non-bare form allomorphs with a bare-to-non-bare ratio 73:157.2 = 1:2.2.

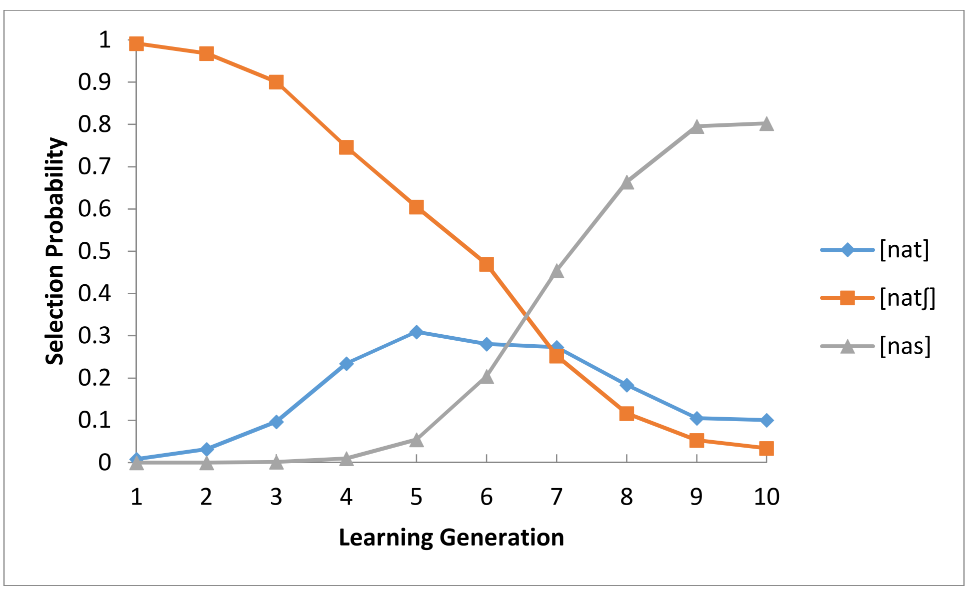


Figure 5.13. SP development of surface allomorphs of 'day' over ten learning generations

The patterns which are not captured in the simulated results are the inter-context variations: The preference for the standard pronunciation in the nominative and locative/dative contexts, but a roughly equal chance for the standard pronunciation and stem-final [s] variants in the accusative context. I will return to discuss this issue in §4.7.

4.6.4.3 Stem: 'flower'

PSI-OT-GLA also generally makes correct predictions of the production of the stem 'flower' in different contexts as illustrated in Table 5.10; the dominant output patterns are still the standard pronunciation with a stem-final aspirated affricate [tʃʰ], and the secondary output is almost always with a stem-final [s] except in the locative/dative context.

'flower'	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[k'otʃʰ-i] 80%	[k'os-i] 20%	[k'otʃʰ-ɨl] 60%	[k'os-ɨl] 40%	[k'otʃʰ-e] 90%	[k'otʰ-e] 10%
Sim	[k'otʃʰ-i] **65.6%**	[k'os-i] **34.1%**	[k'otʃʰ-ɨl] **68.2%**	[k'os-ɨl] **25.8%**	[k'otʃʰ-e] **66.8%**	[k'os-e] **25.2%**

Table 5.10. Percentages of the primary and secondary outputs of 'flower' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)

The SP development in Figure 5.14 indicates that the non-bare form allomorph [k'otʃʰ] is also the primary stem input choice in Generation 5, which leads to high percentages of standard pronunciation outputs in all three contexts. Nevertheless, the stem has a frequency bias toward its bare form allomorph [k'ot] with a bare-to-non-bare ratio 2.2:1. Therefore, the bare form allomorph eventually receives a dominant SP as the primary input choice since Generation 6.

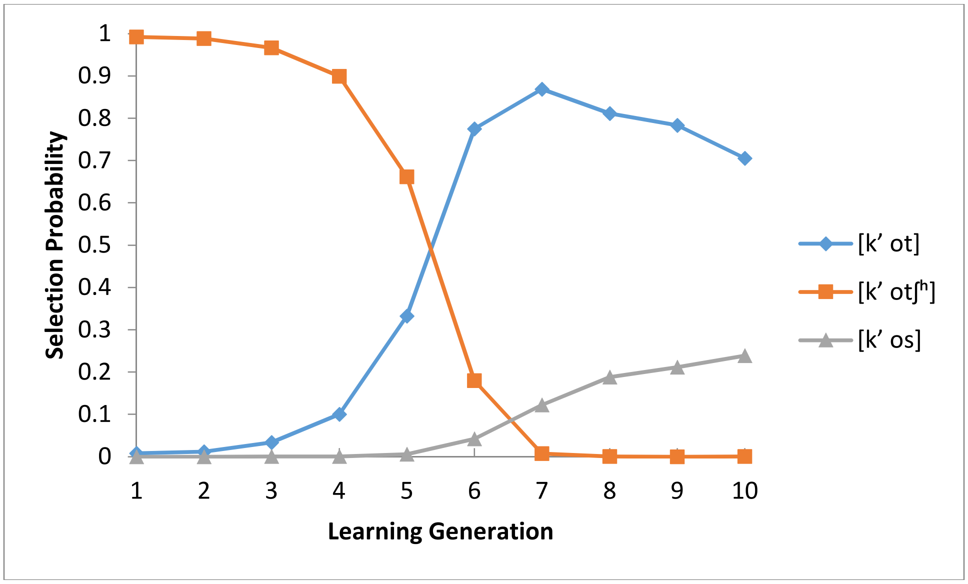


Figure 5.14. SP development of surface allomorphs of 'flower' over ten learning generations

4.6.4.4 Stem: 'wall' and 'rice'

The simulated results of the two stems 'wall' and 'rice' are summarized in Table 5.11, which perfectly matches the performance data in Jun & Lee's experiment with an output pattern of their standard pronunciation across the three contexts without an exception. The results nevertheless come with no surprise. The two stems have only one surface allomorph (i.e. [pjək] and [pap]) that can always be selected as the stem input and thus surfaces as the standard pronunciation without violating any constraints involved in the simulation.

'wall'	nominative -i		accusative -ɨl		locative/dative -e	
Exp	[pjək-i] 100%		[pjək-ɨl] 100%		[pjək-e] 100%	
Sim	[pjək-i] **100%**		[pjək-ɨl] **100%**		[pjək-e] **100%**	
'rice'	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[pap-i] 100%		[pap-ɨl] 100%		[pap-e] 100%	
Sim	[pap-i] **100%**		[pap-ɨl] **100%**		[pap-e] **100%**	

Table 5.11. Percentages of the primary and secondary outputs of 'wall' and 'rice' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs

4.6.5 Mismatches between simulation and experimental result

This section focuses on the stems left out from the discussion in §4.6.4, whose simulated results have some minor or significant mismatches in the comparison with the experimental results. Output patterns of these stems will be discussed below along with possible causes.

4.6.5.1 Stem: 'clothes'

The simulated results of the stem 'clothes' is summarized in Table 5.12, which suggest that PSI-OT-GLA still predicted the standard pronunciation to be the dominant pattern in the three contexts. The mismatches occur as secondary outputs with a stem-final [tʃ] are generated frequently in the simulated results but are completely absent in the experimental results.

'clothes'	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[os-i] 100%		[os-ɨl] 90%	[otʃʰ-ɨl] 10%	[os-e] 100%	
Sim	[os-i] **82.5%**	[otʃ-i] **17.3%**	[os-ɨl] **80.5%**	[otʃ-ɨl] **13.5%**	[os-e] **76.2%**	[otʃ-e] **15.3%**

Table 5.12. Percentages of the primary and secondary outputs of 'clothes' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs

This mismatch can be attributed to a frequency bias toward the bare form allomorph of the stem 'clothes' with a bare-to-non-bare ratio 261:130.95 = 1:0.5. As more standard pronunciations can be derived from the input /ot+Suffix/ via the /t/→[s] mapping, therefore lowering the IEP of [ot] and MEP[X+Ø]$_{\text{BARE}}$, the frequency bias leads to a growing SP(/ot/) as in Figure 5.15. In Generation 5, there is only a chance of 65% to select /os+Suffix/ as the input, which can surface as the standard pronunciation [o.s-Suffix]. However, since IDENT(del rel)-IO does not fully dominate IDENT(cont)-IO, only another 11~17% of the standard pronunciation can be derived via the the /t/→[s] mapping from the input /ot+Suffix/; the rest of the /ot+Suffix/ inputs surface as [otʃ-Suffix] when IDENT(del rel)-IO is probabilistically outranked by IDENT(cont)-IO.

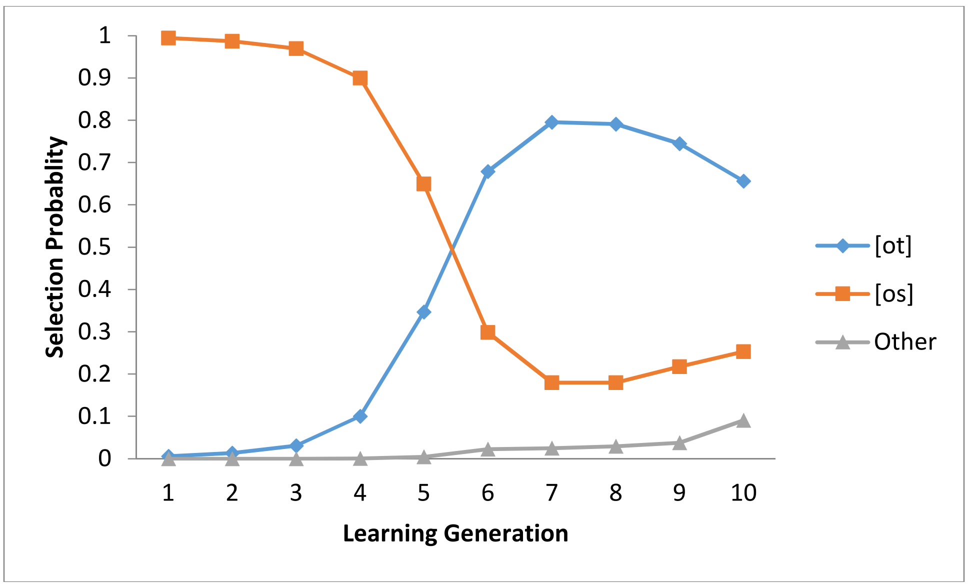


Figure 5.15. SP development of surface allomorphs of 'clothes' over ten learning generations

This mismatch may disappear with either a more accurate frequency counting or a full-scale simulation with a large training corpus. The frequency bias toward the bare form allomorph may not be found existing in more precise frequency estimation, and SP(/ot/) may thus still be dominant to allow more standard pronunciation outputs to be derived from the input /os+Suffix/. A full-scale simulation that includes every Korean noun form in the training corpus might also promote IDENT(del rel)-IO to be much higher than IDENT(cont)-IO with a higher number of /t/→[s] mapping. Recall that there are only two stems with a stem-final

[s] in their suffixed forms, which might not be adequate to build a solid ranking for the /t/→[s] mapping.

4.6.5.2 Stem: ‘breast’

For the stem ‘breast’ in the nominative context, the simulated results produced by PSI-OT-GLA do not seriously diverge from the experimental results as summarized in Table 5.13, but in the accusative context, the primary and secondary outputs are opposite. In the experimental results, the primary output is the stem-final [s] variant, and the secondary output is the standard pronunciation, whereas the standard pronunciation is still the dominant output patterns predicted by PSI-OT-GLA.

‘breast’	nominative *-i*		accusative *-ɨl*		locative/dative *-e*
Exp	[tʃətʃ-i] 50%	[tʃəs-i] 50%	[tʃəs-ɨl] 90%	[tʃətʃ-ɨl] 10%	None
Sim	[tʃətʃ-i] **68.5%**	[tʃəs-i] **31.4%**	[tʃətʃ-ɨl] **65.6%**	[tʃəs-ɨl] **30.2%**	

Table 5.13. Percentages of the primary and secondary outputs of ‘breast’ in Generation 5 of the simulated results (Sim) and Jun & Lee’s experimental results (Exp)

The variation patterns in the simulated results are also attributed to a higher SP(/tʃətʃ/) in Generation 5 as shown in Figure 5.16, which is associated with an the extreme frequency bias toward non-bare form allomorphs with a bare-to-non-bare ratio 12:153 = 1:12.8. The standard pronunciation [tʃətʃ+Suffix] can thus surface frequently from the input /tʃətʃ+Suffix/. This dominant SP(/tʃətʃ/) does not change based on different morphosyntactic contexts, and thus cannot capture the inter-context variations, which will be discussed in §4.7.

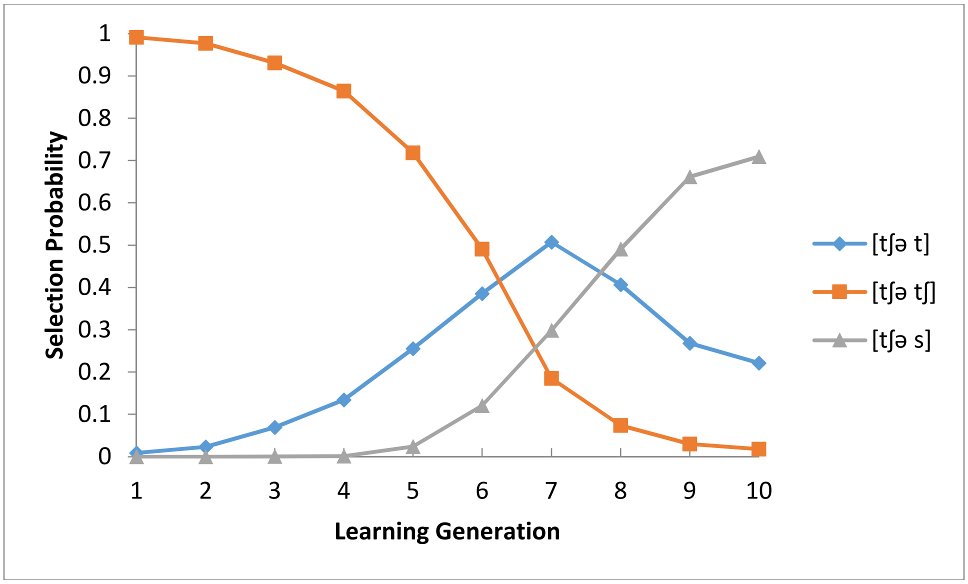


Figure 5.16. SP development of surface allomorphs of 'breast' over ten learning generations

4.6.5.3 Stem: 'face'

When simulating the production of the stem 'face' in the three contexts, PSI-OT-GLA, although correctly making the prediction that standard pronunciation is more prevalent, clearly overestimates the frequency of the standard pronunciation (Table 5.14). The reason is that with a slightly unbalance distribution in favor of the non-bare form allomorph /natʃʰ/ (bare-to-non-bare ratio 6:9.55 = 1:1.6), the non-bare form allomorph is always considered as the best stem input option from the very first generation (see Figure 5.17).

'face'	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[natʃʰ-i] 70%	[nas-i] 30%	[natʃʰ-ɨl] 60%	[nas-ɨl] 40%	[natʃʰ-e] 54.5%	[nas-e] 45.5%
Sim	[natʃʰ-i] **99.9%**	[nas-i] **0.1%**	[natʃʰ-ɨl] **99.6%**	[natʰ-ɨl] **0.3%**	[natʃʰ-e] **99.6%**	[nas-e] **0.4%**

Table 5.14. Percentages of the primary and secondary outputs of 'face' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)

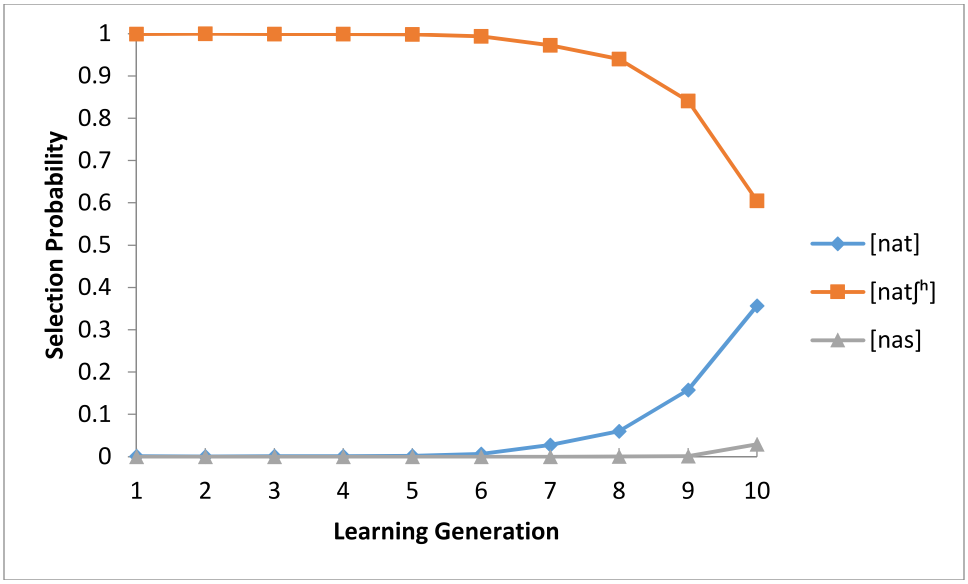


Figure 5.17. SP development of surface allomorphs of 'face' over ten learning generations

This stem, however, has a relatively lower frequency count in the Sejong text corpus. It is thus possible that in a real speech corpus the stem is produced as the bare form more often than in the text corpus. If there is indeed a frequency bias toward the bare form allomorph [nat], the stem-final [s] variants will emerge from the input /nat+Suffix/ via the /t/→[s] mapping to produce variation patterns similar to those in the experimental results.

4.6.5.4 Stem: 'field'

The first significant mismatch exists in the simulated results of the stem 'field' summarized in Table 5.15. In the experiment data, the primary output is always the stem-final [tʃʰ] variant in the nominative and accusative contexts, but in the same contexts in the simulated results, the primary output is the stem-final [s] variant. The reason is that when the non-bare form allomorph /patʰ/ or the bare form allomorph /pat/ is selected as the stem input in the two contexts with a higher SP (see Figure 5.18), the stem-final consonant cannot surface faithfully to avoid violating higher-ranked *Ti or *t/_]$_{\text{STEM}}$-S. However, since IDENT(cont)-IO has a slightly lower constraint value, a stem-final [tʰ] or [t] has a higher chance to surface as a stem-final [s].

'field'	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[patʃʰ-i] 90%	[pas-i] 10%	[patʃʰ-ɨl] 80%	[pas-ɨl] 20%	[patʰ-e] 90%	[patʃʰ-e] 10%
Sim	[pas-i] **52.3%**	[patʃʰ-i] **27.1%**	[pas-i] **46.7%**	[patʃʰ-i] **16.7%**	[patʰ-e] **44.7%**	[pas-e] **33.2%**

Table 5.15. Percentages of the primary and secondary outputs of 'field' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)

In the locative/dative context, the experimental results indicate the prevalence of the standard pronunciation with a stem-final [tʰ], which, albeit being greatly underestimated, is captured in the simulated results. The higher proportion of the standard pronunciation can be attributed to the lack of higher-ranked constraints forcing any alternation before the locative/dative context, as predicted in §3.3. The underestimation originates in the higher $SP_{/pat/}$ that drives the /t/→[s] mapping for the stem-final [s] variants, which are absent in the experimental results.

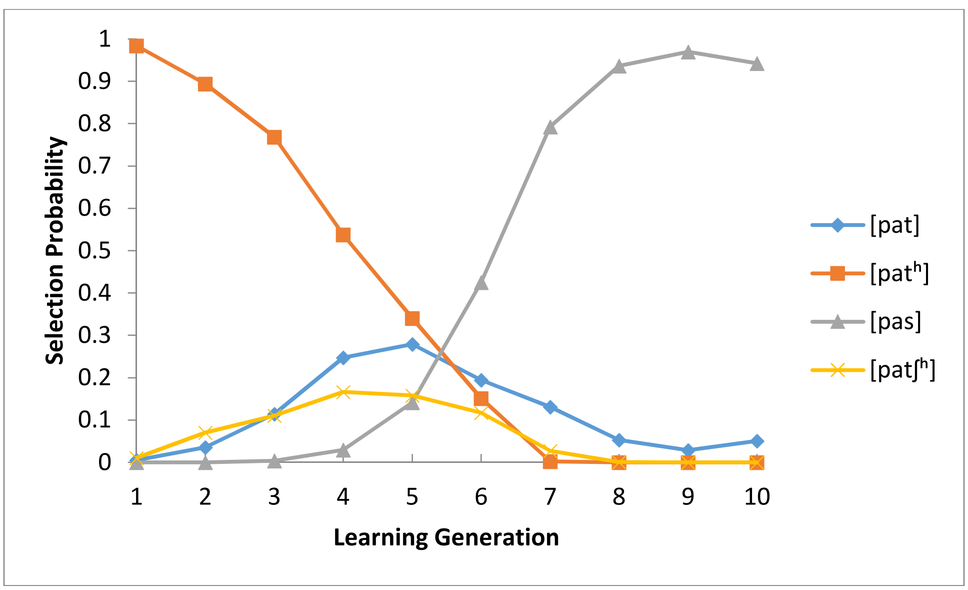


Figure 5.18. SP development of surface allomorphs of 'field' over ten learning generations

4.6.5.5 Stem: 'red bean'

The simulated results of the stem 'red bean' resemble the experimental results in terms of the similar proportions of stem-final [s] variants, which is the consequence of a high SP$_{/p^hat/}$ (see Figure 5.19) and the grammar for the /t/→[s] mapping.[71] The considerable mismatch that undermines the simulation of this stem is the significant proportion of stem-final [tʃ] variants, which are impossible in the experimental results. Recall that the /t/→[tʃʰ] mapping is harmonically-bounded by the /t/→[tʃ] mapping with the constraints involved in the current simulation (see §3.1); selecting the bare form allomorph /pʰat/ thus can only result in stem-final [tʃ] variants.

'red bean'	nominative -*i*		accusative -*ɨl*		locative/dative -*e*	
Exp	[pʰatʃʰ-i] 54.5%	[pʰas-i] 44.5%	[pʰas-ɨl] 60%	[pʰatʃʰ-ɨl] 30%	[pʰas-e] 45.5%	[pʰatʰ-e] 36.4%
Sim	[pʰas-i] **51.3%**	[pʰatʃ-i] **48.7%**	[pʰas-ɨl] **43.6%**	[pʰatʃ-i] **37.5%**	[pʰatʃ-e] **39.5%**	[pʰas-e] **36.6%**

Table 5.16. Percentages of the primary and secondary outputs of 'red bean' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp)

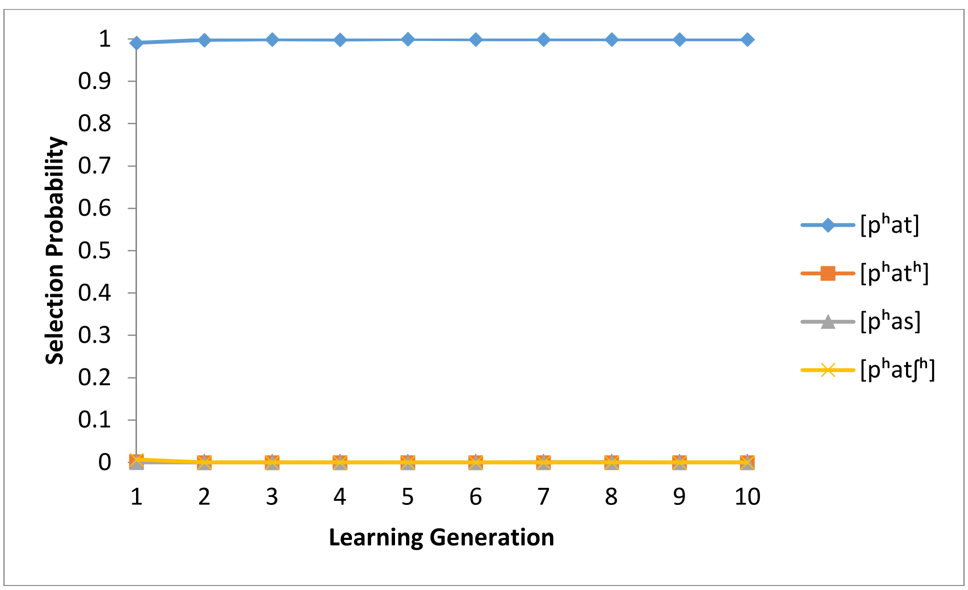


Figure 5.19. SP development of surface allomorphs of 'red bean' over ten learning generations

Two adjustments are possible to improve the current simulated results. The stem 'red bean' is again a low frequency stem, which almost never appears in any

[71] Based on the frequency count after correction, the stem 'red bean' almost never appears in any suffixal context and the extremely low token frequencies (0.4 and 0.2 in the nominative and accusative contexts) and can be easily erased from the lexicon due to the memory decay effect.

suffixal contexts in the text corpus count. An extreme frequency bias toward the bare form allomorph of the stem thus occurs in the training corpus of this simulation. By examining spontaneous speech corpora, we might find a different frequency distribution which raises the SP of non-bare form allomorphs like [pʰatʰ]. A stem-final [tʃʰ] can thus be derived in the nominative and accusative contexts from [tʰ]. Another line is to switch to an Output-Output constraint framework, which I will return to in §4.7.

4.6.5.6 Stem: 'kitchen', 'outside', and 'leaf'

Some mismatches also appear in the simulated results of stems with a stem-final labial/velar stop, including 'kitchen', 'outside', and 'leaf', which will be discussed in this section. Let us first consider the results of 'kitchen' in Table 5.17: The simulated results generally approximate the experimental results with the dominant devoiced surface variations in all three contexts due to the top-ranked unnatural markedness constraint forbidding any stem-final aspirated labial/velar stop. Nevertheless, the secondary output in the simulated results (i.e. standard pronunciation) is not possible in the experimental results.

'kitchen'	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[puək-i] 80%	[puəkʰ-i] 20%	[puək-ɨl] 80%	[puəkʰ-ɨl] 20%	[puək-e] 60%	[puəkʰ-e] 40%
Sim	[puək-i] **100%**		[puək-ɨl] **100%**		[puək-e] **100%**	

Table 5.17. Percentages of the primary and secondary outputs of 'kitchen' in Generation 5 of the simulated results (Sim) and Jun & Lee's experimental results (Exp); grey cells represent the absence of secondary outputs

More significant mismatches can be found in the results of 'outside' and 'leaf' in Table 5.18; the two stems were almost always produced as their standard pronunciation with a stem-final tense or aspirated stop in Jun & Lee's experiment, whereas the stem-final stops are forced to surface as plain stops to satisfy top-ranked unnatural markedness constraints in the simulation.

‘outside’	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[pak’-i] 100%		[pak’-ɨl] 100%		[pak’-e] 60%	
Sim	[pak-i] **100%**		[pak-ɨl] **100%**		[pak-e] **100%**	
‘leaf’	nominative *-i*		accusative *-ɨl*		locative/dative *-e*	
Exp	[ip^{h}-i] 100%		[ip^{h}-ɨl] 70%	[ip-ɨl] 30%	[ip^{h}-e] 100%	
Sim	[ip-i] **100%**		[ip-ɨl] **100%**		[ip-e] **100%**	

Table 5.18. Percentages of the primary and secondary outputs of ‘outside’ and ‘leaf’ in Generation 5 of the simulated results (Sim) and Jun & Lee’s experimental results (Exp); grey cells represent the absence of secondary outputs

The standard pronunciation of these stems, however, is not completely impossible in PSI-OT-GLA. Recall that in early stages that IDENT(asp)-IO and IDENT(tense)-IO can be promoted over the two unnatural markedness constraints as appeared in Generation 1 in §4.6.2, and that the target outputs can be produced in this phase. For example, the stem ‘outside’ has a strong frequency bias toward the non-bare form allomorph [pak’] with a bare-to-non-bare ratio 68:373.85 = 1:5.75. A dominant SP of [pak’] before Generation 4 in Figure 5.20 allows the input like /pak’+i/, and IDENT(tense)-IO will be promoted in order to produce the target output [pa.k’-i]. The issue, then, is how to generate these standard pronunciations with other surface variations at the same stage.

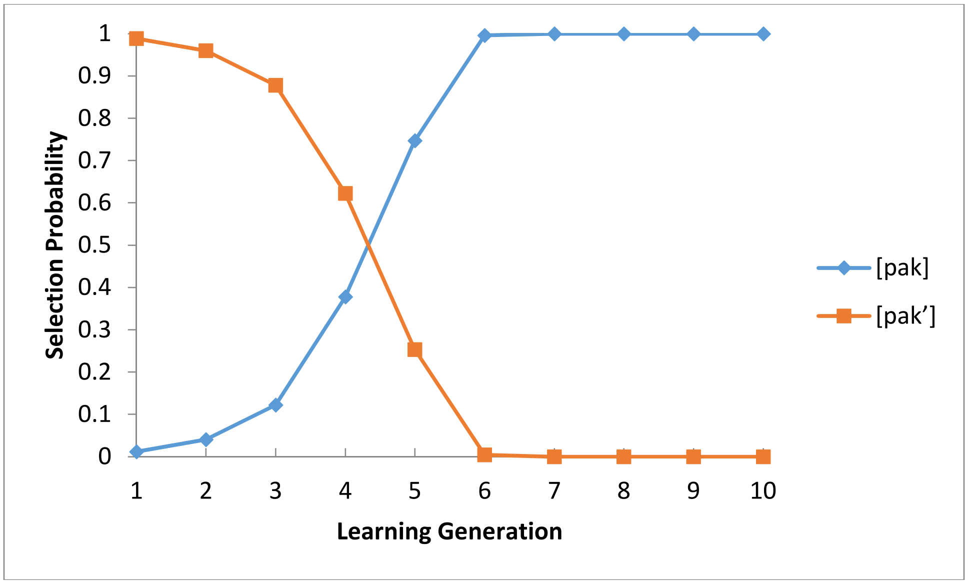


Figure 5.20. SP development of surface allomorphs of 'outside' over ten learning generations

4.7 Discussion

The above comparisons suggest that the simulated results produced by PSI-OT-GLA partially overlap with the experimental results and capture some prediction made earlier in §3. First, for most noun stems the standard pronunciation is still more prevalent as a result of a more dominant SP of non-bare form allomorphs in Generation 5. Second, if the secondary output is possible, it is usually the stem-final [s] variant due to a slightly higher constraint value of IDENT(cont)-IO if compared to that of IDENT(del rel)-IO. Third, the standard pronunciation with a stem-final [t^h] is more likely to be preserved in the locative/dative context without any higher-ranked markedness constraint forcing [t^h] to alternate before the suffix [-e].

The simulated results, albeit seemingly acceptable, are by no means satisfactory unless the issues brought up in §4.6.5 can be solved for further improvements. Particularly, different training corpora containing different information can largely change the simulated results. The current simulated results with mismatches, especially those of the stems with a lower token frequency in the text corpus, might suffer from inaccurate estimation of token frequencies in real speech contexts. The simulated results can also be undermined if the selected items in the training corpus are not representative. The current training corpus includes only thirteen stems in five contexts – a set of 65 Korean noun forms, and there are only two stems with a stem-final [s] in their standard suffixed forms, which are not completely dominant stem-final patterns as revealed by the type frequency count. It is thus expected that a full-scale simulation with all Korean noun forms in the training corpus can reduce the discrepancies found in the foregoing sections. It has also been shown that PSI-OT-GLA cannot capture inter-context variations shown in the experiment data, particularly those in the nominative and accusative contexts, which seem relatively random. However, the smaller number of data points in Jun & Lee's experiment may not be sufficient to verify such variation patterns. Perhaps, inter-context variations in production will disappear after collecting more responses (cf. well-formedness judgement in Jun 2010). Finally, the /t/→[tʃ^h] mapping can be made possible with an modification based on Output-Output Correspondence. For example, assuming the /pat+i/ is the input, the output [pa.tʃ^h-i] is possible when IDENT(asp)-OO is top-ranked to require the [asp] specification in the output to be identical to [+asp] in another suffixed form [pa.t^h-e] as illustrated in Tableau 5.21 below.

/pat+i/ Loc./Dat.: [pa.t^h-e]	IDENT(asp)-OO	*Ti	IDENT(asp)-IO
☞[pa.tʃ^h-i]			*
[pa.t^h-i]		*!	
[pa.s-i]	*!		

Tableau 5.21. Generating stem-final [tʃ^h] variants via OO Correspondence

Do's (2012) experiments of investigating how Korean-learning children acquiring noun paradigms also highlight the necessity of OO Correspondence as the results show the frequent occurrence of stem-final [t] undergoing intervocalic voicing in children's suffixed nouns in the age of 4;2-5;8 (e.g. [pat]-[pa.d-i]-[pa.d-ɨl]); that is, the suffixed forms are leveled to their bare form paradigm. This finding also coincides with Tessier's (2006) proposal that OO constraints are initially top-ranked, and the lack of adult variations patterns during this phase suggests unnatural markedness constraints to be either initially dominated by OO constraints or invented much later at the M stage to dominate OO constraints. Further improvements in the current PSI-OT-GLA model with the integration of OO correspondence will be explored in Ch. 6.

4.8 Advantages over base identification approach

Before closing the chapter, I would like to demonstrate why PSI-OT-GLA, with some significant deviations in its simulated results, still better accounts for the variation patterns if compared to Albright's (2002, 2008, in press) base identification approach adopted previously by Jun & Lee (2007) and Jun (2010), which has been briefly reviewed at the end of §2.

In the base identification approach, all surface allomorphs of the same morpheme are grouped by their morphosyntactic contexts with a set of rules that can derive one form from another. Assuming that there are two pairs of surface forms [pat]/[tʃʌt] and [patʰ-ɨl]/[tʃʌs-ɨl] representing two stems in different morphosyntactic contexts (bare and accusative), a set of rules will be generated for the forms in one context to derive the forms in the other context as in (21) via Minimal Generalized Learners. If the number of correctly derived surface forms (i.e. reliability, in Albright's term) is higher for the rules /t+ɨl/→[tʰɨl] and /t+ɨl/→[sɨl], learners consider the paradigms [pat] and [tʃʌt] in bare context to be the base of [patʰɨl] and [tʃʌsɨl]; all accusative forms are thus derived from their bare forms, and the weight of the two rules is determined by its type frequency. Since the rule /t+ɨl/→[sɨl] has a higher type frequency, it has a high weight and thus is applied more frequently to derive stem-final [s] variants.

(21) Learning via base identification

bare		acc.
[pat]	←——→	[patʰɨl]
[tʃʌt]	←——→	[tʃʌsɨl]
/t+ɨl/→[tʰɨl]		/tʰɨl/→[t-Ø]
/t+ɨl/→[sɨl]		/sɨl/→[t-Ø]

Based on the above example in (21), however, if frequency information is excluded, accusative forms should always be recognized as the bases since deriving base forms from accusative forms coincides with the coda neutralization analysis, and thus the two rules /tʰɨl/→[t-Ø] and /sɨl/→[t-Ø] do not generate any exception.[72] Albright (2008) proposes that the effect of token frequency can be accounted for in the base identification approach as the reliability of a set of morphophonological

[72] Accusative paradigms in Korean, when serving as the base, may in fact derive surface errors as the accusative suffix has multiple surface allomorphs – [-ɨl] for consonant-ending stems and [-ɾɨl] for vowel-ending stems. Since stem-final /l/ also changes to [ɾ] before [-ɨl] the [ɾ-ɨl] and [-ɾɨl] sequences create a parsing ambiguity, which lead to errors in the base identification approach. Since the issue raised by this parsing ambiguity does not change the conclusion of the following discussion, it is ignored at this point.

rules is underestimated using Mikheev's (1997) confidence score function: The less frequent a set of rules are tested, the more severely their reliability is underestimated. For example, if two sets of rules never derive any output errors but are tested 100 (i.e. 100/100) and five times (i.e. 5/5) respectively, the corresponding corrected reliabilities are 0.99 and 0.825.[73] That is, for rules that are tested less frequently, learners should be less 'confident' with the robustness of these rules (see §1.2 of Ch. 2 and Jarosz (2011) for a similar effect of 'test frequency' in morphophonological acquisition).

In the Korean case, since accusative forms are much rare, learners are less confident with the rules that derive other paradigms from accusative forms, despite the possibility that these rules can generate more correct outputs. Assuming that bare forms and the corresponding rules can generate 50 correct outputs out of 75 tests and accusative forms can generate four out of five tests (i.e. tested with the same frequency ratio 75:5 reported in Lee's 1999 study), the raw reliability is higher for accusative forms (50 / 75 = 0.667 vs. 4 / 5 = 0.8). However, the corrected reliability instead shows higher confidence in bare forms and their rules (0.628 vs. 0.607), allowing bare forms to be identified as the base.[74]

While this base identity approach can also explain how the base is chosen by referring to token frequency as well, token frequency is not directly involved in determining the base. It could be the case that the rules /t+ɨl/→[tʰɨl] and /t+ɨl/→[sɨl] are tested more frequently but still end up having a lower corrected reliability due to a great amount of output errors they derive. Furthermore, it misses individual generalizations as individual token frequencies of different paradigms from the same context accumulate as a whole to illustrate only the frequency difference between different morphological classes, and the base selection is always categorical; either one or another set of paradigms are the base for all paradigms, and there is no item-specific variation. As summarized in §4.1, token frequency can vary depending on the nature of individual lexical items; the noun 'outside', for example, is naturally produced more frequently in the locative/dative context than other noun stems.

---

[73] The calculation of the corrected reliabilities in Albright's modeling slightly differs from that in Mikheev (1997); the former calculates confidence interval from a t-distribution and the latter from a normal distribution (Adam Albright, p.c., Dec. 2011).

[74] As in PSI-OT-GLA, base identification do not always predict the paradigm set with a highest token frequency to be the base as shown by the modeling of Korean verb paradigm variations in Albright & Kang (in press).

I suggest that these individual variations in the frequency distribution must contribute to the surface variations of different suffixed noun forms as indicated in Jun & Lee's results. For example, in the experimental results of the stem 'field', with a higher token frequency of its locative/dative form [pat^h-e] (62 in the text corpus), the allomorph /pat^h/ should be selected to construct all surface forms, leading to the dominance of [pat^h-e] (loc./dat.; 90%), [patʃ^h-ɨl] (acc.; 80%) and [patʃ^h-i] (nom.; 90%) via assibilatory aspiration. On the other hand, the stem 'red bean', which seems to have almost no suffixed form in the text corpus, must build its suffixed form from the bare form allomorph /p^{h}at/ rather than the allomorph /p^{h}at^h/ from its standard locative/dative form [p^{h}at^h-e]. The /t/→[s] mapping then generates a greater number of stem-final [s] variants in the experimental results: [p^{h}as-e] (loc./dat.; 45.5%), [p^{h}as-ɨl] (acc.; 60%), and [p^{h}as-i] (nom.; 44.5%). In sum, although the two stems have a stem-final [t^h] in their standard locative/dative suffix forms, their surface variation patterns diverge significantly due to their different frequency distribution. Various factors may result in some noticeable disparities in its simulated results as discussed in §4.7, but PSI-OT-GLA should be considered to be more flexible than base identification with the possibility of predicting above individual variations.

5. Local summary

The stem-final variations in Korean appear to be a great challenge for PSI-OT-GLA, which require the model to generate similar patterns with multiple stem allomorphs and stem-final mappings. Since the simulated results are by no means perfect, it is easy to cast doubt on the precision of the current version of PSI-OT-GLA. Having said that, the simulated results should have sufficiently demonstrated the importance of appreciating and incorporating individual differences, such as different distributional frequencies, among different stems in different contexts, which become the advantages of adopting PSI-OT-GLA in modeling the Korean stem-final variations. To further improve from this privilege, a future expansion of PSI-OT-GLA will be sketched in Ch. 6 as a strong commitment to reduce the gap between experimental and simulated results in terms of learning morphophonological alternations and developing morphophonemic diachronic changes.

Chapter 6

Toward an Output-Output correspondence model

The simulated results produced by PSI-OT-GLA were shown to resemble the results found in the experiments eliciting the production of morphophonemic alternations in the previous chapters. This final chapter will focus on how to further improve the precision of the morphophonological learning model.

The similarity and major mismatches between experimental and simulated results are summarized in §1. A revised Output-Output (OO) correspondence model is then proposed in §2 as an extension to reduce the gap between the two results. In this alternative model, learners do not probabilistically selects the input of a morphologically complex form from a set of surface allomorphs, and the input of the form is always identical to its target output. For example, the input of the Dutch word [bɛ.d-ən] 'beds' is always /bɛd+ən/. The influences from morphologically-related or unrelated paradigms instead lie in the correspondence between their outputs; e.g. the output of 'beds' might be *[bɛ.t-ən] when a faithful mapping with the singular form [bɛt] is required. This revised model will be applied in §3 to account for the patterns which the original PSI-OT-GLA is too limited to generate in the Dutch, Mandarin, and Korean simulations. In particular, a small-scale simulation demonstrates how the model acquires the Dutch morphophonology. In §4, I illustrate the primary challenges to the assumption in the original PSI-OT-GLA that only surface-true allomorphs can be possible phonological inputs. In languages like English, Palauan, and Tonkawa, a non-surface true input must be created to derive all surface forms, which is a collection of the possible segments of its surface allomorphs and is thus referred to as 'superset UR'. Without this superset UR, PSI-OT-GLA cannot acquire the target grammar and may generate some surface variations that are unlikely to occur by randomly selecting surface allomorphs as the input. Phonologically conditioned allomorph selection such as in Polish also requires input listing, which is also not possible in PSI-OT-GLA. The newly proposed OO model will solve these potential problems since every morphologically-related form has its own input, and learners' only have to acquire a grammar that can produce an output faithfully; a single superset UR for all surface forms and input listing are thus unnecessary.

1. Success and failure of PSI-OT-GLA

Simulations in previous chapters demonstrate the capability of PSI-OT-GLA in modeling temporary and permanent morphophonemic changes triggered by relexicalization which, as a result of a frequency bias toward a set of surface paradigms, leads to grammar shifts. In Ch. 3, the frequency bias toward singular allomorphs of alternating stems prompts learners to select singular allomorphs as the stem input in the production of plural forms (e.g. /bɛt+ən/, not /bɛd+ən/ as the input for [bɛ.d-ən] 'beds'). Consequently, a great number of devoicing output errors such as *[bɛ.t-ən] are generated as observed in the experimental results, which can be acknowledged as the primary success of this simulation work. However, to derive the correct output from different inputs, PSI-OT-GLA predicts that *V[-voi]V is promoted to temporarily outrank IDENT(voi)-IO until learners successfully recognize plural allomorphs as better input options, and this re-ranking process at times gives rise to intervocalic voicing output errors of non-alternating plural forms (e.g. /pɛt+ən/→*[pɛ.d-ən] 'caps'). While such output errors are attested in the experimental results, PSI-OT-GLA clearly overestimates the number of these errors.

The simulation of the Mandarin Tone 3 change from a full concave tone to a shortened low-dipping tone in Ch. 4 also successfully predicts that Tone 3 words occurring in non-final positions undergo the change faster with a stronger frequency bias toward their low-dipping tone allomorph, and that the constraint *MLH is eventually promoted to be top-ranked and bans all full concave tones. The same tendency appeared in the experimental results as well, with a lower proportion of the full concave tone produced with Tone 3 words occurring more frequently in non-final positions. The proportions in the simulated results also generally coincide with those in the experimental results. However, for a small number of Tone 3 words with an extreme frequency bias toward their low-dipping tone allomorph, PSI-OT-GLA significantly underestimates the number of full concave tone outputs.

PSI-OT-GLA also succeeded in modeling the initialization of stem-final variations in Korean suffixed noun forms in Ch. 5, which were assumed to be triggered by a general frequency bias toward bare form allomorphs. As bare form allomorphs gradually become the primary stem input options, various stem-final obstruents in suffixed forms must be derived from a stem-final plain stop /p/, /t/, and /k/ in the bare form allomorphs. Stem-final variations thus occur as a result of deriving different stem-final obstruents in the suffixed forms from input plain stops. As in previous studies, stem-final /s/ variations are the most prevalent variation pattern in the simulated results of PSI-OT-GLA due to a higher number of /t/→[s] mappings that demote IDENT(cont)-IO lower than other faithfulness constraints. PSI-OT-GLA nonetheless was unable to generate different preferred suffixed form variants across suffixal contexts, and incapable of mapping an input stem-final /t/ to an aspirated affricate [tʃʰ] for highly productive stem-final [tʃʰ] variants.

Considering only where PSI-OT-GLA succeeded altogether, the model satisfies the four criteria for designing a formal learning model in §5 of Ch. 1. First of all, PSI-OT-GLA has formal sufficiency, as the model can clearly explain how a learning outcome can be generated with a training corpus via a sequence of transparent learning processes (i.e. constraint re-ranking and adjustments of lexical parameters) in all three simulations. With the same set of natural and unnatural constraints and lexical parameters for adults and children in the simulation of Dutch, Mandarin and Korean, the criterion explanatory continuity is also satisfied. The temporary overgeneralization patterns captured in the Dutch simulation also demonstrated developmental compatibility that children may undergo successive stages during morphophonological acquisition. Finally, divergent acceptability was demonstrated in Ch. 4 and 5 in which diachronic changes were predicted as the model accepted 'imperfect' learning outcomes which could not generate all possible surface forms correctly due to an unbalanced distribution of learning inputs.

The mismatches between simulation and experimental results above, however, require some patches to improve the existing model. As suggested in the discussion in previous chapters, an Output-Output (OO) correspondence model seems a plausible solution. The reason that the original PSI-OT-GLA did not incorporate OO correspondence for a full lexical network is twofold. First, without a full lexical network, the simulation design as well as the constraint set can be simplified to examine the rich lexicon hypothesis in morphophonological acquisition with less computational effort. Secondly, paradigmatic leveling patterns in morphophonological acquisition can be partially captured in this simplified design without OO correspondence. As demonstrated in Ch. 2 and 3, for example, when IDENT(voi)-IO is ranked higher than *V[-voi]V to forbid intervocalic voicing and the stem input of an alternating plural form (e.g. 'beds') is the singular allomorph (e.g. /bɛt+ən/), the output error *[bɛ.t-ən] (cf. [bɛ.d-ən]) appears to be leveled to its singular form [bɛt] as repeated in Tableau 6.1.

/bɛt+ən/ Singular: [bɛt]	IDENT(voi)-IO	*V[-voi]V
✓bɛ.t-ən		*
☹bɛ.d-ən	*!	

Tableau 6.1. Paradigm leveling via input selection with higher ranked faithfulness constraints

In sum, PSI-OT-GLA can be treated as a more basic model that provides a preliminary test of the rich lexicon hypothesis with a simpler design but remains ambitious in simulating real performance data. Its similar learning outcomes, as demonstrated from Ch. 3 to Ch. 5, then offer the basis of building a computationally more sophisticated rich lexicon model with OO correspondence to further approximate real morphophonological learning. The goal of this chapter is thus to illustrate the transition from the input selection process to OO correspondence and explain how the OO model will help reduce the discrepancy between simulated results and real production data.

2. A rich lexicon and Output-Output correspondence

As claimed in Ch. 1, PSI is not introduced as the only or universal strategy that learners could implement with a rich lexicon in morphophonological acquisition, and some may have already considered the PSI framework too conservative in modeling a rich lexicon since it only addresses the association between morphologically related paradigms. To expand the current model to capture complex associations in a rich lexicon and solve the issues encountered in the previous case studies by PSI-OT-GLA, it is time to explore the possibility of integrating Output-Output (OO) correspondence with the proposed framework.

OO correspondence allows constraint-based phonology to build the association between lexical entries by examining the identity between derived forms and their underived stems as initially proposed in Benua (1995, 1997). McCarthy (2005) extended the original OO framework to multi-directional mappings between all morphologically-related paradigms and each paradigm can serve as the reference for paradigmatic leveling (see §3.4 of Ch. 1). Myers (2002) also seeks to formalize analogy via OO constraint conjunction constraints such as in (1). For example, OO(drive,dive:[ayv])^OO($drive_{past}$,$dive_{past}$:[o]) requires the past tense of 'drive' and 'dive' to agree in the nucleus [o] since their present tense 'drive' and 'dive' also agree in their rhyming segment [ayv]. OO correspondence thus seems not only appropriate to depict morphological relationships in the lexicon as in its initial design but also suitable for establishing a complete lexical network to explore lexical effects in generative phonology.

(1) OO(*a*,*c*:F)^OO(*b*,*d*:G): If the forms *a* and *c* agree in a set of features F, the forms *b* and *d* must also agree in a set of feature G.

This section serves as a rudimentary sketch of such an OO correspondence model and its application in morphophonological acquisition. As an expansion of the original PSI-OT-GLA, the OO model still seeks to incorporate the same lexical factors to determine basic paradigms in different ways, hopefully inheriting the advantages in modeling morphophonological acquisition from its predecessor.

In §2.1, paradigms with lexical associations in phonotactic learning and morphophonological acquisition will be defined. Crucially, morphologically related paradigms are only associated with each other in the latter. The formal definition of OO constraints in the proposed OO framework is introduced in §2.2: Each paradigm has a corresponding OO constraint which requires a faithful mapping between the output and a paradigm. An unfaithful mapping between the output and a high-frequency paradigm incurs more violations of the paradigm's OO constraints and is thus less preferred; i.e. Paradigms with a higher token frequency are more likely to be selected as a morphological base and the reference for paradigmatic leveling as in PSI-OT-GLA. The evaluation of the chance for each output candidate to be optimal in the OO model will be spelled out in §2.3. One major change in this evaluation process from PSI-OT-GLA is to adopt Harmonic Grammar to constantly evaluate the accumulative token frequency effect of neighbor paradigms represented by a set of OO constraints. The initial constraint weight of each constraint and the weight adjustment process along the acquisition course will be explained in §2.4, which is followed by a brief discussion of other possible influences on constraint weights in §2.5.

2.1 Lexical entry, lexical paradigm, and lexical associations

The construction of the current OO model starts with the definition of a lexical entry and lexical paradigm. The former can be construed as a 'label' of a set of phonetic forms that share same semantic definition in the lexicon. For example, the lexical entry of the phonetic form [bɛd-ən] in Dutch has the semantic definition of 'bed (noun, plural)'. A lexical paradigm, on the other hand, represents a member, or one single token, of a lexical entry, which is stored faithfully in the lexicon (i.e. Lexicon Optimization) via an auditory channel.[75] Each two lexical paradigms of a lexical entry are then associated together and form a lexical network as in Figure 6.1, which is essentially identical to the concept of 'exemplar cloud' in Exemplar-based Theory (e.g. Pierrehumbert 2000, 2002; Bybee 2001). The output of a lexical entry is not derived from any specific phonological input as every paradigm of the entry can influence the output via OO correspondence as shown in §2.2 later.

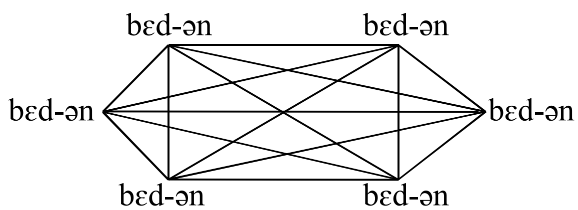


Figure 6.1. Lexical association between six [bɛd-ən] tokens

Associations between lexical paradigms can also be extended to link morphologically related paradigm pairs (i.e. 'word-to-word associations'; Burzio 2002, 2005). For instance, the paradigms of the plural form [bɛd-ən] can potentially be associated with their singular form paradigms [bɛt-ən] as shown in Figure 6.2. In other words, the output of a lexical entry is not only influenced by its own lexical paradigms but all the associated paradigms of other lexical entries. Nevertheless, I assume that these associations are much weaker in phonotactic learning (P stage) than in morphophonological acquisition (M stage). The dotted lines in Figure 6.2, for instance, stand for the associations that are strengthened only after morphological awareness in the M stage (also see §2.5 for lexical associations emerged from semantic similarity).

[75] Since lexical paradigms must be stored via perception, the lexicon only includes surface-true paradigms and thus abstract lexical representations cannot be referred to as a morphological base in an OO correspondence; the OO model, like PSI-OT-GLA, thus obeys the Single Surface Base Restriction proposed in Albright (2002, 2008).

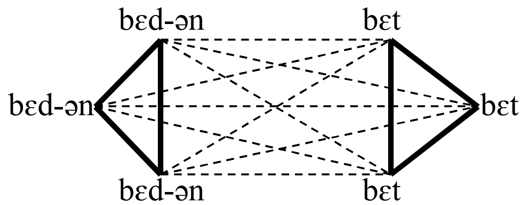


Figure 6.2. Lexical associations between [bɛt] and [bɛd-ən] tokens in the P stage (solid lines) and the M stage (dotted lines)

2.2 Lexically-indexed Output-Output correspondence

In a constraint-based grammar, the influence on the output of a lexical entry from lexical paradigms (i.e. outputs stored from the production of the same or different speaker) can be formalized in an OO correspondence model. The core assumption of the revised OO framework is that learners are allowed to create a set of lexically-indexed OO (LIOO) constraints for each lexical entry, which is an extension of Pater's (2000, 2008) original proposal. Violations of a LIOO constraint are the pressure to force the uniformity between an output and every paradigm member of the lexical entry indexed by the constraint. After introducing LIOO constraints, IO correspondence becomes redundant since the faithful production of a word form is forced by the pressure of leveling to all of its lexical paradigms via OO correspondence, rather than to any specific lexical input via IO correspondence. For example, the output [bɛd-ən] is produced to preserve the faithful OO mapping between the output and each of its six stored paradigms in Figure 6.1. As a consequence, this OO model will not include IO faithfulness constraints. It should also be noted that the input selection process in the original PSI model can be fully abandoned in the OO model since the phonological input needs not to be determined first to produce an output faithful to a certain paradigm via IO correspondence; the faithful mapping between an output and a lexical paradigm is directly expressed with OO correspondence.

An example of LIOO constraints that requires faithful feature mappings can be formally defined in (2); IDENT-OO(bɛd-ən; [voi]) is violated when an output and a token (i.e. lexical paradigm) of [bɛd-ən] do not agree in the specification of the feature [voi].

(2) IDENT-Output-Output(P;F): The output and each paradigm P of a lexical category must agree in the features F.

Tableau 6.2 illustrates an example of the violation of a LIOO constraint with the production of [bɛ.d-ən], which was assumed to have six tokens in the lexicon as in Figure 6.1 in the previous section. If the intervocalic voiced stop is devoiced in the output, IDENT-OO(bɛd-ən;[voi]) is violated six times since the faithful mapping between the output and each token of [bɛ.d-ən] is examined by the LIOO constraint. Such a constraint definition can also predict that analogical changes are less likely to be extended to words with a higher token frequency (Pierrehumbert 2001, 2002; Bybee 2001); by accepting analogical changes, the output of high-frequency paradigms incurs more LIOO constraint violations.

/bɛd-ən/	IDENT-OO(bɛd-ən;[voi])
☞bɛ.d-ən	
bɛ.t-ən	******

Tableau 6.2. Violation of a LIOO constraint indexed to the paradigm [bɛd-ən]

When morphologically related paradigms are associated in the lexical network in the M stage (see Figure 6.2 in §2.1), their LIOO constraints jointly evaluate the correspondence between the output and the two sets of the paradigms. Assuming that the LIOO constraints are unranked, the token frequency of each entry determines the basic paradigm (i.e. which paradigms the output should be leveled to): The output will be leveled to a paradigm with a higher token frequency to avoid a higher number of LIOO constraint violations. Therefore, as in PSI-OT-GLA, token frequency serves as a lexical factor to select a morphological base. In addition, when the token frequency effect is an intrinsic property of constraint violation, there is no need to invent a complex function to convert token frequencies into constraint values as required in the lexical constraint framework (see §6.3 of Ch. 2).

With the pair [bɛt]~[bɛd-ən], for example, two LIOO constraints IDENT-OO(bɛt ; [voi]) and IDENT-OO(bɛd-ən;[voi]) can be proposed. If the two entries [bɛt] and [bɛd-ən] have ten and five lexical paradigms respectively, the output will be leveled to the former in the output as shown in Tableau 6.3. By contrast, when the form [bɛd-ən] has a higher token frequency in the lexicon, the intervocalic voicing contrast will be preserved in the output as in Tableau 6.4.

/bɛd-ən/	IDENT-OO(bɛd-ən;[voi])	IDENT-OO(bɛt;[voi])
bɛ.d-ən		******!****
☞bɛ.t-ən	*****	

Tableau 6.3. Paradigmatic leveling due to a higher frequency of [bɛt]

/bɛd-ən/	IDENT-OO(bɛd-ən;[voi])	IDENT-OO(bɛt;[voi])
☞bɛ.d-ən		*****
bɛ.t-ən	******!****	

Tableau 6.4. Faithful output due to a higher frequency of [bɛd-ən]

The interaction between markedness and LIOO constraints may block paradigmatic leveling. As in Tableau 6.5, the output of the singular form [bɛt] cannot be leveled to the plural form [bɛ.d-ən] even if [bɛ.d-ən] has a higher token frequency, due to a top-ranked *VOICEDOBSCODA (i.e. voiced obstruent codas are prohibited). Or, as in Tableau 6.6, when *V[-voi]V (i.e. no intervocalic voiceless segments) dominates the two LIOO constraints, the output of the plural form is still [bɛd-ən] despite a higher token frequency of the singular form [bɛt].

/bɛt/	*VOICEDOBSCODA	IDENT-OO (bɛd-ən;[voi])	IDENT-OO (bɛt;[voi])
☞bɛt		**********	
bɛd	*!		*****

Tableau 6.5. Paradigmatic leveling of singular forms is prohibited due to top-ranked *VOICEDOBSCODA

/bɛ.d-ən/	*V[-voi]V	IDENT-OO (bɛd-ən;[voi])	IDENT-OO (bɛt;[voi])
☞bɛ.d-ən			**********
bɛ.t-ən	*!	*****	

Tableau 6.6. Paradigmatic leveling of plural forms is prohibited due to higher-ranked *V[-voi]V

The above examples demonstrated how the source of paradigmatic leveling can be determined upon different token frequencies in the OO model as in PSI-OT-GLA. The primary difference is that in PSI-OT-GLA, paradigmatic leveling is not directly related to token frequency; token frequency is only responsible for selecting different phonological inputs like /bɛt+ən/ vs. /bɛd+ən/, and the ranking of IO-Faithfulness constraints determines whether paradigmatic leveling occurs when /bɛt+ən/ is the input. The OO model merges the two seemingly independent processes (i.e. input selection and IO-Faithfulness evaluation) into one single evaluation process with LIOO constraints.

2.3 LIOO in Harmonic Grammar

Although the token frequency effect in the OO model can be captured with different violation numbers, LIOO constraints must be unranked in OT for violation counts to take effect due to strict domination: Violations of lower-ranked constraints cannot cancel violations of higher-ranked constraints. Thus, if an LIOO constraint indexed to a low-frequency lexical entry strictly dominates another LIOO constraint indexed to a high-frequency lexical entry for some reason (see §2.4), the output must always be faithful to the paradigms of the low-frequency entry as in Tableau 6.7 even if there is a significant frequency difference between the two entries.

/bɛd-ən/ Sg: bɛt	IDENT-OO(bɛd-ən;[voi])	IDENT-OO(bɛt;[voi])
☞bɛ.d-ən		***********************
bɛ.t-ən	*!	

Tableau 6.7. Token frequency differences obscured by strict domination

This is not saying that low-frequency entries can never be the source of paradigmatic leveling, but that language models should always take the token frequency of each entry into consideration at any given point, so an extreme frequency bias toward an entry can always prompt language learners to level an output to the paradigms of the entry as in the case of Mandarin and Korean diachronic changes captured by PSI-OT-GLA in Ch. 4 and 5. To this end, the proposed OO model adopts Harmonic Grammar (HG; Legendre et al. 1990, 2006), in which constraints are no longer ranked with strict domination as in OT. Instead, each constraint has a negative constraint weight, and the 'score' of each output candidate is the sum of each constraint weight multiplied by the violation number of each constraint as defined in (3). Among a set of output candidates, the one with the highest score (i.e. closest to zero) is the winner.

(3) The score of an output candidate h(x) in HG can be defined as follows, where w_i is the constraint weight of the i^{th} constraint of *N* constraints, and Ci(x) is the number the i^{th} constraint being violated by *x*. (Hayes & Wilson 2008)

$$h(x) = \sum_{i=i}^{N} w_i C_i(x)$$

Each score can also be converted into an output probability in a Maximal Entropy model to model surface variation patterns. As in Hayes & Wilson (2008), the raw score is first exponential-transformed into a maxent value as in (4); the lower is the raw score, the smaller is the number after the transformation. Each exponential-transformed score is then divided by the sum of all exponential-transformed scores at hand to calculate the probability for each output candidate to be the winner as in (5).

(4) A maxent value P*(x) of an output candidate *x* can be defined as follows, where h(x) represents the raw score of *x.*

$$P*(x) = \exp(-h(x))$$

(5) An output probability P(x) of an output candidate *x* can be defined as follows, where Ω represents a set of all possible output candidates, and P*(y) represents the maxent value of an output candidate *y*, which is a member of Ω.

$$P(x) = \frac{P*(x)}{\sum_{y \in \Omega} P*(y)}$$

In this HG-based model, since each violation mark of a LIOO constraint constantly contributes to the calculation of output scores and probabilities, the frequency effect is always active. Meanwhile, although candidates violating constraints with a more negative constraint weight are more likely to be ruled out, it is also possible that more violations of constraints with a less negative weight can produce a lower score, which is dubbed as a 'gang effect'. An LIOO constraint with a lower weight in HG can still force paradigmatic leveling if there is an extreme frequency bias toward the entry indexed by the constraint.

As in Tableau 6.8, the weight of IDENT-OO(bɛd-ən;[voi]) is much higher than IDENT-OO(bɛt;[voi]) (-1 vs. -0.1), suggesting a stronger pressure for an output to be faithful to /bɛd-ən/ paradigms. Nevertheless, if /bɛt/ has a much higher token frequency, say 23, any output unfaithful to /bɛt/ paradigms violates IDENT-OO(bɛt;[voi]) 23 times. The score of the output candidate [bɛ.d-ən], which is unfaithful to [bɛt] paradigms, thus turns out to be lower ($-0.1 \times 23 = -2.3$) as there is presumably one single /bɛd-ən/ token which the output candidate [bɛ.t-ən] is unfaithful to (i.e. output score = $-1 \times 1 = -1$). In sum, even though the candidate [bɛ.d-ən] satisfies the higher-weighed constraint IDENT-OO(bɛd-ən;[voi]), the chance for the output candidate [bɛ.t-ən] to win is actually higher.

		IDENT-OO(bɛd-ən;[voi])	IDENT-OO(bɛt;[voi])
Weight		-1	-0.1
bɛ.d-ən	-2.3		***********************
☞bɛ.t-ən	-1	*	

Tableau 6.8. Token frequency bias capture in HG with a gang effect

An additional advantage of adopting the HG framework in this OO model (henceforth LIOO-HG) is to evaluate the influence of the entire association network. Possible lexical associations are not limited to the different tokens of the same or morphologically related forms but can be extended to link every possible lexical item together, and each association can be examined with a LIOO constraint indexed to a different paradigm (see §2.5 below). The evaluation process in HG considers all violations to LIOO constraints and can thus model every possible leveling pressure in the lexical network.

2.4 Initial state of constraints and weight adjustment in LIOO-HG

As in PSI-OT-GLA, the completion of morphophonological acquisition in LIOO-HG includes the transformation from an initial state to an end state of the constraint grammar. This section serves to define the initial state of constraints in LIOO-HG and the process of adjusting constraint weights gradually during the acquisition course.

Markedness and LIOO constraints are assumed to have a weight of -100 and 0 respectively by default to be in accord with the original proposal that markedness constraints are initially more dominant than faithfulness constraints to force unmarked output patterns at the beginning of phonological acquisition. To complete phonotactic learning, learners must gradually promote LIOO constraints and demote markedness constraints to produce marked phonological structures and elements in a target language. As a consequence, some LIOO constraints must form a strong faithful pressure by the end of the P stage, which may in turn lead to paradigmatic leveling at the beginning of the M stage, which is captured by positing initially top-ranked OO constraints in Jesney & Tessier (2011) and Tessier (2007). Tableau 6.9 illustrates the weight adjustment process of IDENT-OO(pɛt-ən; [voi]) and *V[-voi]V when an output is wrongly produced with intervocalic voicing as in PSI-OT-GLA.

/pɛt-ən/	*V[-voi]V	IDENT-OO(pɛt-ən;[voi])
Weight	-100	0
✓pɛt-ən -100	*→	
☞☹pɛd-ən -1		←*

Tableau 6.9. Weight adjustment triggered by output errors in HG; '✓' = correct output, '☹' = output error, '→' = demotion, '←' = promotion

During constraint promotion and demotion, it is also necessary to implement a biased weight adjustment algorithm to ensure restrictive learning in LIOO-HG, which is defined in (6). That is, while the weight of markedness constraints is adjusted by the original plasticity, the weight of faithfulness constraints is adjusted by a much smaller amount.

(6) Weight adjustment bias in LIOO-HG can be defined as follows, where r represents raw plasticity, r_{Mk} represents the plasticity of markedness constraints, r_{LIOO} represents the plasticity of LIOO constraints, and n_{LIOO} and n_{Mk} represent the number of LIOO and markedness constraints respectively. (Jesney & Tessier 2011:269)

$$r_{Mk} = r \qquad r_{LIOO} = \frac{r_{Mk}}{n_{LIOO} \times n_{Mk}}$$

This weight adjustment process parallels the function of Individual Error Proportion (IEP) introduced in PSI-OT-GLA as an attempt to cancel a frequency bias toward an allomorph that can lead to more output errors. For example, in Tableau 6.10, the singular form entry [bɛt] has a higher token frequency, and paradigmatic leveling occurs to avoid the large violation number of IDENT-

OO(bɛt;[voi]). The weight adjustment process then seeks to further increase the gap between the two LIOO constraints and thus to cancel the frequency bias toward the singular paradigm. In other words, LIOO-HG does not have to record the number of output errors as in PSI-OT-GLA to compute basic paradigms; the number of output errors is reflected in the weight adjustment per se. As discussed in Ch. 2, however, it is possible that not every frequency bias can be cancelled at the end of morphophonological learning. For reasons of space, the threshold for learners to recover from a frequency bias in LIOO-HG is not discussed here.

/bɛd-ən/	IDENT-OO(bɛd-ən;[voi])	IDENT-OO(bɛt;[voi])
Weight	-1	-0.1
✓bɛ.d-ən -2.3		***********************→
☞☹bɛ.t-ən -1	←*	

Tableau 6.10. Constraint promotion and demotion triggered by output errors in HG; '→' = demotion, '←' = promotion

2.5 Semantic similarity and OO correspondence in LIOO-HG

The above sections only focused on the OO correspondence between an output and its semantically identical or morphologically related and thus semantically highly similar paradigms. We nevertheless also have to consider weaker associations created between lexical paradigms of entries that only share a few semantic features. In the current LIOO-HG, LIOO constraints indexed to these entries will generate an equal leveling pressure between an output and the paradigms of these entries despite their semantic differences. While it is expected that all associated lexical paradigms can join together to affect the production of an output, those who are semantically less similar to the output should be less likely to make a strong impact. For example, the two Dutch words 'dog (pl.)' [hɔnd-ən] and 'turtle (pl.)' [sxilpɑd-ən] are possibly associated in the lexicon by sharing some semantic features of animals, but it will be surprising if one of the 'animal paradigms' is leveled to another. Summarizing, the pressure of paradigmatic leveling between an output and a set of lexical paradigms is directly related to the semantic similarity between them, which should be formalized in LIOO-HG to avoid overgeneralizing leveling patterns.

One way to quantify the leveling pressure based on the semantic similarity between lexical entries is to have LIOO constraints indexed to various (sub)sets of semantic properties other than those indexed to specific lexical entries. Therefore, the more semantic properties are shared by an output and a lexical paradigm, the more LIOO constraints are violated if the output is not faithful to the lexical paradigm (i.e. a stronger pressure to level the output with the lexical paradigm). To illustrate how this approach works, consider a mini-lexicon composed of three lexical entries [hɔnd-ən] 'dog (pl.)', [hɔnt] 'dog (sg.)', and [sxilpɑt] 'turtle (sg.)', and each of them has only one token. We can then assume three LIOO constraints indexed to the entry 'dog (pl.)', all entries sharing the root morpheme 'dog', and all entries sharing the [+animal] feature as in (7).

(7) LIOO constraints indexed to different semantic features

a. OO('dog (pl.)';F): The output and each paradigm of 'dog (pl.)' must agree in the features F.
b. OO(root='dog';F): The output and each paradigm with the root morpheme 'dog' must agree in the features F.
c. OO([+animal];F): The output and each paradigm with the semantic feature [+animal] must agree in the features F.

In the production of 'dog (pl.)', three possible scenarios are represented with corresponding output candidates in Tableau 6.11. The output of 'dog (pl.)' and its semantically and phonetically identical lexical paradigms of 'dog (pl.)' [hɔnd-ən] share all the three semantic properties (i.e. 'dog (pl.)', root='dog', [+animal]). Therefore, the output must be faithful to the paradigms of 'dog (pl.)' for the fewest violation marks; the stem of the output [hɔnd-ən] disagree with [hɔnt] in one segment and [sxilpɑt] in all seven segments. When the root is leveled to the paradigms of 'dog (sg.)' as in [hɔnt-ən], all the three constraints are violated due to an unfaithful mapping with the paradigms of 'dog (pl.)'; the output is thus slightly less preferred, but the similar violation number demonstrates a nearly equal chance for the output to be leveled to the singular form of 'dog' as predicted by the high semantic similarity between 'dog (pl.)' and 'dog (sg.)'. Finally, when the output of 'dog (pl.)' is completely leveled to [sxilpɑt] and surfaces as [sxilpɑt-ən], it draws extra violations not only because it fails to be faithful to all the paradigms sharing the root morpheme 'dog' (i.e. [hɔnd-ən] and [hɔnt]) but also because it largely deviates from these paradigms with many segmental mismatches; a low phonetic similarity further weakens the pressure to level the output of 'dog (pl.)' to [sxilpɑt].[76]

[76] See also Burzio (2002, 2005) and Myers (2002) as well for the discussion of the effect phonetic

‘dog (pl.)’	OO (‘dog (pl.)’;F)	OO (root=‘dog’;F)	OO ([+animal];F)
☞[hɔnd-ən]		*[hɔnt]	*[hɔnt], *******[sxilpɑt]
[hɔnt-ən]	*	*[hɔnd-ən]	*[hɔnd-ən], *******[sxilpɑt]
[sxilpɑt-ən]	*******	******[hɔnt], *******[hɔnd-ən]	******[hɔnt], *******[hɔnd-ən]

Tableau 6.11. Semantic similarity between outputs and lexical paradigms as leveling pressure

An alternative, which will be adopted for its simpler implementation in the simulation discussed later, is a weight adjustment function of Evaluator (or EVAL) in HG, which manipulates raw constraint weights based on the semantic similarity between an output and a lexical paradigm; the lower the semantic similarity between an output and a paradigm, the more the weight of the LIOO constraint indexed to the paradigm decreases in an output evaluation process. In the simulation in §3.1, I will assume that the raw weight of the LIOO constraint indexed to the paradigms semantically identical to the output remains unchanged, whereas the raw weight of the LIOO constraints indexed to the paradigms morphologically related to the output is divided by 1.1 and the weight of remaining LIOO constraints is divided by 1.2. The setting is obviously arbitrary and oversimplified, but with more details in either of the above implementations to be specified, which are beyond the scope of the chapter, the dividing function serves as a plausible choice to directly add the effect of semantic similarity in the simulation in §3.1 below.

3. Potential improvements in previous simulations with LIOO-HG

After building the basic structure of LIOO-HG, I will discuss how possibly the gaps between simulation and experimental results summarized in §1 could be filled by the OO model in this section in order of Dutch, Mandarin, and Korean. In particular, a small scale simulation of the Dutch case based on LIOO-HG will demonstrate a learning outcome more similar to real learners’ performance.

similarity on paradigmatic leveling and analogy.

3.1 Dutch final devoicing

The improvements in LIOO-HG is firstly demonstrated with a small scale simulation of the acquisition of Dutch final devoicing via Java® programming,[77] in which the training corpus only includes the singular and plural forms of one non-alternating and alternating stem (foot and bed) in Table 6.1. As in Ch. 3, learners randomly perceive one of the four forms based on the distributional probabilities and randomly select one form to produce in every learning cycle.

Stem	Sg.	Freq.	Prob.	Pl.	Freq.	Prob.
foot	[vut]	96	0.184	[vut-ən]	129	0.248
bed	[bɛt]	284	0.545	[bɛd-ən]	12	0.023

Table 6.1. Mini-corpus with raw token frequencies in CELEX (Baayen et al. 1995) and corresponding distributional probabilities

The constraint set in this simulation includes the same five markedness constraints listed in (8), which have an initial constraint weight of -100 as defined in §2.4. IO-Faithfulness constraints IDENT-IO, MAX-IO, and DEP-IO in are replaced with corresponding LIOO constraint types in (9). Since each of the three constraint types is indexed to all four lexical paradigms, there will be twelve LIOO constraints in total in this small-scale simulation with an initial constraint weight of zero. The biased error-driven weight adjustment process follows the description in §2.4, and the upper bound is set to zero to maintain non-positive constraint weights. Further adjustments are made to raw constraint weights in the output evaluation process based on semantic similarity as mentioned at the end of §2.5.

(8) Markedness constraints in small-scale simulation

a. ONSET: Syllables without onset are prohibited.
b. *VOICEDOBS: Voiced obstruents are prohibited.
c. *VOICEDOBSCODA: Voiced obstruent codas are prohibited.
d. *CODA: Syllable codas are prohibited.
e. *V[-voi]V: Intervocalic voiceless consonants are prohibited.

[77] The source code is available at http://hdl.handle.net/10402/era.39162.

(9) LIOO constraints in small-scale simulation

a. IDENT-OO(*P*,F): The output and every token of a lexical paradigm *P* must agree in the features F.

b. MAX-OO(*P*): The segments in every token of a lexical paradigm *P* must have a correspondent in the output.

c. DEP-OO(*P*): The segments in the output must have a correspondent in every token of a lexical paradigm *P*.

As elaborated in §2.1 and §2.2, in the P stage, morphologically related paradigms are not associated with each other, and therefore an output only violates LIOO constraints indexed to the paradigm which the output represents, as shown in Tableau 6.12. After the association between morphologically related paradigms are built in the M stage, the faithful mapping between an output and all associated paradigms will be evaluated via various LIOO constraints indexed to the paradigms (Tableau 6.13).

/bɛd-ən/	IDENT-OO(bɛt;[voi])	IDENT-OO(bɛd-ən;[voi])
☞✓bɛ.d-ən		
bɛ.t-ən		*!

Tableau 6.12. No violation of other LIOO constraints in the P stage; '✓' = correct output

/bɛd-ən/	IDENT-OO(bɛt; [voi])	IDENT-OO(bɛd-ən; [voi])
✓bɛ.d-ən	*!	
☞☹bɛ.t-ən		*

Tableau 6.13. Violation(s) of LIOO constraints indexed to morphologically related paradigms in the M stage; '✓' = correct output, '☹' = output error

The outcome of this simulation is expected to demonstrate (i) the intervocalic devoicing errors of alternating plural forms like *[bɛ.t-ən] due to a frequency bias toward the singular paradigm, (ii) the recovery process from this frequency bias in which the plural paradigm is recognized as the morphological base at the end of the M stage simulation, and (iii) an overall low error rate of the non-alternating plural form, which yet slightly rises along with the decrease in the error rate of the alternating plural form in the middle of the learning course. The first two patterns are identical to the simulated results in Ch. 3, and the third pattern is the potential improvement in LIOO-HG after the introduction of between-paradigm mappings.

The number of learning cycles is 10,000 and 100,000 in the two stages, and below I will start the discussion of the simulated results with the constraint weights at the end of the P stage simulation. As shown in Tableau 6.14, I only include the constraints crucial to the acquisition of final devoicing. The total absence of voiced obstruent codas results in the unchanged high weight of *VOICEDOBSCODA. Although at this point, *V[-voi]V has a higher weight than IDENT-OO(vut-ən;[voi]), it does not mean that intervocalic voicing is wrongly acquired in LIOO-HG due to gang effect as illustrated below.

Constraint	*VOICEDOBSCODA	*V[-voi]V	IDENT-OO (vut-ən;[voi])
Weight	-100	-8	-6.7

Tableau 6.14. The weight of constraints crucial to final devoicing

The accumulated token number of [vut-ən] is 2,430 after 10,000 learning cycles in the P stage simulation, which is similar to the expected number based on the distributional probability in Table 6.1 (i.e. 10,000 × 0.248 = 2,480). Recall that the number of violations of a paradigm's LIOO constraint should be the token frequency of the paradigm per se. However, the product of a high token frequency and a constraint weight may lead to a large negative raw score which, after exponential-transformation, becomes an extremely small positive number eventually treated as zero in the computer simulation; the probability of output candidates cannot be calculated in this case. To solve this problem, each accumulated raw token frequency is log-transformed beforehand to avoid a large negative raw score of output candidates. As in Tableau 6.15, the constraint weight of the LIOO constraint is multiplied with the log-transformed token number of [vut-ən], and the score of the output candidate [vud-ən] becomes much lower than that of [vut-ən]. Summarizing, after phonotactic learning, LIOO-HG successfully acquires a grammar that forbids both voiced obstruent codas and the intervocalic voicing process.

	*V[-voi]V	IDENT-OO(vut-ən;[voi])
Weight	-8	-6.7
☞✓[vut-ən] -8	*	
[vud-ən] -52		log(2,534)≈7.8

Tableau 6.15. Intervocalic voicing blocked by gang effect; '✓' = correct output

After morphologically related paradigms are associated via a lexical network in the M stage, output errors of plural forms should emerge either when learners may initially recognize the singular paradigms as morphological bases and level plural forms toward the singular paradigms, or when learners overgeneralize the intervocalic voicing pattern in the production of non-alternating plural forms. To study the error patterns, an elicitation task is simulated for every 100 learning cycles in the M stage simulation as in Ch. 3, in which LIOO-HG produces the two plural forms for 100 times with the constraint grammar at the given point. The devoicing error rates of [bɛ.d-ən] and the intervocalic voicing error rates of [vut-ən] from the 100^{th} learning cycle to the $10{,}100^{th}$ learning cycles are plotted in Figure 6.3.

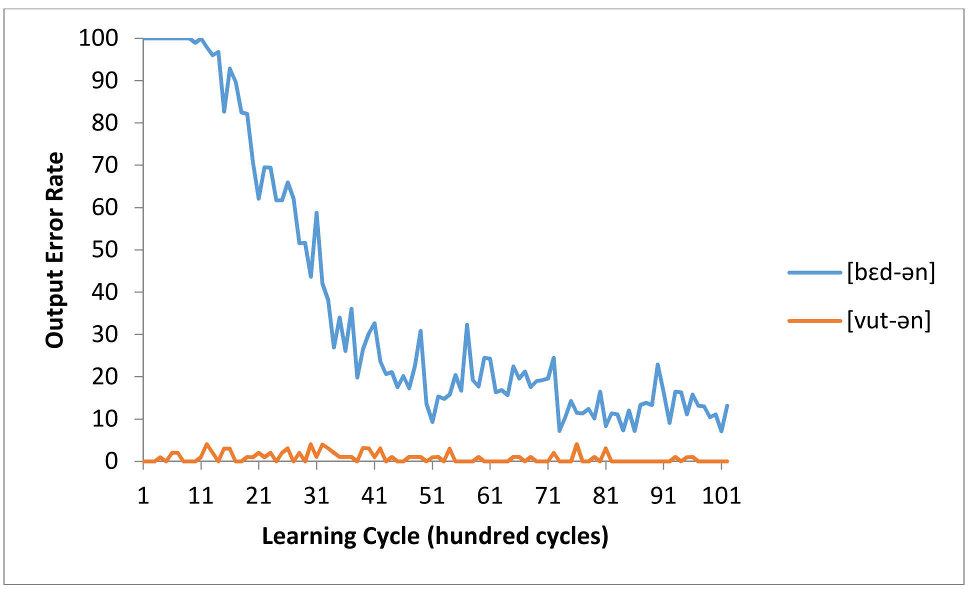


Figure 6.3. Output error rates of [bɛd-ən] and [vut-ən] from the 100^{th} learning cycle to the $10{,}100^{th}$ learning cycle in the M stage simulation

The results indicate that the error rate of [bɛ.d-ən] is much higher at the beginning of the M stage simulation, which is not surprising because of the extreme frequency bias toward the singular paradigm [bɛt] (see Table 6.1). That is, the bias allows the token frequency of [bɛt] to accumulate faster to give rise to a great violation number of IDENT-OO(bɛt;[voi]) and thus significantly lower the score of any output candidate unfaithful to [bɛt], including [bɛ.d-ən]. Furthermore, with a higher weight of IDENT-OO(bɛt;[voi]) after the semantic similarity adjustment as illustrated in Tableau 6.16, the score difference is significant for the output error *[bɛ.t-ən] to be always selected as the winner at this point. To fix this, the error-driven weight adjustment process increases the weight of *V[-voi]V and IDENT-OO(bɛd-ən; [voi]) to raise the output probability of [bɛ.d-ən] whenever the output error *[bɛ.t-ən] is produced.

'beds'		IDENT-OO (bɛt;[voi])	*V[-voi]V	IDENT-OO (bɛd-ən;[voi])
Weight		-14.3 / 1.1 = -13	-8	-4.3
✓[bɛ.d-ən]	-111.8	log(5,556)≈8.6→		
☞☹[bɛ.t-ən]	-31.3		←*	←log(225)≈5.4

Tableau 6.16. Output error *[bɛ.t-ən] emerges with a higher output score; constraint weights and accumulated token frequencies in the 100^{th} learning cycle of the M stage (i.e. $10{,}100^{th}$ learning cycle from the beginning of the simulation); '✓' = correct output, '☹' = output error, '←' = promotion, '→' = demotion

Note that the improvement in the error rate of [bɛ.d-ən] is coupled with the slightly rising error rate of [vut-ən] between the $1{,}100^{th}$ learning cycle and the $4{,}100^{th}$ learning cycle, which was found in Kerkhoff's (2007) experimental results and previously captured by PSI-OT-GLA. The reason is that whenever the output error *[bɛ.t-ən] is produced, the weight of *V[-voi]V also slightly increases to help preserve the intervocalic voiced obstruent as demonstrated above. This adjustment thus raises the probability of the overgeneralized intervocalic voicing pattern causing non-alternating plural errors like *[vud-ən].

Unlike in PSI-OT-GLA, however, such an error pattern is much rarer in LIOO-HG, which is more consistent with the experimental results. The sharp difference can be ascribed to the pressure to preserve the identity between paradigms of non-alternating stems in LIOO-HG, which is absent in PSI-OT-GLA. In the pair [vut]~[vut-ən], the voicing specification of the stem-final obstruent is identical, and the application of intervocalic voicing in the output candidate [vud-ən] is doubly penalized by IDENT-OO(vut;[voi]) and IDENT-OO(vut-ən;[voi]). Therefore, despite of an increasing weight of *V[-voi]V, the low score of the output candidate [vud-ən] suggests a low productivity of the overgeneralization pattern as demonstrated in Tableau 6.17.

'caps'	*V[-voi]V	IDENT-OO (vut-ən;[voi])	IDENT-OO (vut;[voi])
Weight	-28	-7.1	-0.4 / 1.1 = -0.36
☞✓[vut-ən] -28	*		
[vud-ən] -58.8		log(2,664)≈7.9	log(2,067)≈7.6

Tableau 6.17. Low score of output error *[vud-ən] with constraint weights and accumulated token frequencies in the 1,100th learning cycle of the M stage simulation; '✓' = correct output

Finally, after the 10,000th learning cycle in the M stage simulation, the plural paradigm [bɛ.d-ən] overcomes the frequency deficit and is recognize as the morphological base in LIOO-HG as IDENT-OO(bɛd-ən; [voi]) gains a higher weight than IDENT-OO(bɛt; [voi]), and the correct output becomes the dominant pattern as demonstrated in Tableau 6.18. Following this development is the disappearance of the rare overgeneralization pattern due to the lack of motive to increase the weight of *V[-voi]V.

'beds'	*V[-voi]V	IDENT-OO (bɛd-ən;[voi])	IDENT-OO (bɛt;[voi])
Weight	-47	-10.7	-10.2 / 1.1 = -9.27
☞✓[bɛd-ən] -86.3			log(11,043)≈9.3
[bɛt-ən] -112.7	*	log(464)≈6.1	

Tableau 6.18. Dominant correct output pattern [bɛd-ən] with constraint weights and token frequencies in the 10,000th learning cycle of the M stage simulation; '✓' = correct output

To sum up here, the implementation of LIOO-HG not only parallels PSI-OT-GLA in generating a similar learning outcome but further approximates the experimental results with the pressure of paradigm uniformity. While there are still some sophisticated issues that need to be addressed before a full-scale simulation is possible, one should be optimistic with its capability of modeling morphophonological acquisition. In the forthcoming sections, I will further spell out the potential improvements in LIOO-HG in the Mandarin and Korean case.

3.2 Mandarin Tone 3 reduction

The uniformity between paradigms required in LIOO-HG may also account for the experimental results in Ch. 4, in which non-final-preferring Tone 3 words were still frequently produced with a full concave tone, i.e. the Tone 3 of these words has not been completely reduced to a low-falling tone (or 'half-tone'). The same Tone 3 words, nevertheless, were predicted to always surface with a half-tone by PSI-OT-GLA due to an extreme frequency bias toward their half-tone allomorph. At the end of Ch. 4, I suggested a possible solution to this major mismatch to be a lexical network that associates all Tone 3 words together; the pressure of paradigmatic uniformity created by the lexical association could thus force non-final-preferring Tone 3 words to be leveled to final-preferring Tone 3 and produced with a full concave tone.

This potential solution is naturally compatible with LIOO-HG and is sketched with an example below. Assume a Mandarin noun 黨 [tɑ̃ŋ^{213}] 'political party' and transitive verb 擋 [tɑ̃ŋ^{213}] 'block'. Given the half-tone and full-tone allomorphs of each Tone 3 words, four corresponding LIOO constraints can be proposed in (10). A noun has a higher chance to be produced in a phrase-final position, which means a higher token frequency of the full-tone allomorph (MLH) of 黨 'political party'. Conversely, since a transitive verb must precede an object, 擋 'block' must be produced in a non-final position more frequently for a higher token frequency of its half-tone allomorph (ML). The output of 擋 'block' is thus expected to be leveled to its half-tone allomorph [tɑ̃ŋ^{21}] to avoid a higher violation number of IDENT-OO(tɑ̃ŋ$^{21}{}_{block}$;T). That is, if we do not consider the association between different Tone 3 words, non-final-preferring Tone 3 words like 擋 'block' should surface mostly with a half-tone to as predicted in PSI-OT-GLA (Tableau 6.19).

(10) LIOO constraints for [tɑ̃ŋ^{213}] 'political party' and [tɑ̃ŋ^{213}] 'block'

a. IDENT-OO(tɑ̃ŋ$^{213}{}_{party}$; T): The output and every token of [tɑ̃ŋ^{213}] 'political party' must agree in the tonal specification.
b. IDENT-OO(tɑ̃ŋ$^{21}{}_{party}$; T): The output and every token of [tɑ̃ŋ^{21}] 'political party' must agree in the tonal specification.
c. IDENT-OO(tɑ̃ŋ$^{213}{}_{block}$; T): The output and every token of [tɑ̃ŋ^{213}] 'block' must agree in the tonal specification.
d. IDENT-OO(tɑ̃ŋ$^{21}{}_{block}$; T): The output and every token of [tɑ̃ŋ^{21}] 'block' must agree in the tonal specification.

擋 'block'	IDENT-OO(tãŋ$^{213}_{block}$;T)	IDENT-OO(tãŋ$^{21}_{block}$;T)
Weight	-1	-1
☞tãŋ^{21} -1	*	
tãŋ^{213} -5		*****

Tableau 6.19. Paradigmatic leveling toward the half-tone verb paradigm

The strong associations between the two Tone 3 words nevertheless put the output under scrutiny of all the four LIOO constraints in (10). The higher token frequency of the full-tone allomorph of the noun 黨 'political party', thus a potential high violation number of IDENT-OO(tãŋ$^{213}_{party}$;T), then creates an opposite force that requires the output of 擋 'block' to be leveled to the paradigm [tãŋ^{213}] of 黨 'political party'. Accordingly, as demonstrated in Tableau 6.20, the two output candidates [tãŋ^{21}] and [tãŋ^{213}] of 擋 'block' may have a similar raw score in LIOO-HG, and an output with a full tone is possible for non-final-preferring Tone 3 words, which agrees with the experimental results in Ch. 4.

擋 'block'	IDENT-OO (tãŋ$^{213}_{block}$; T)	IDENT-OO (tãŋ$^{21}_{block}$; T)	IDENT-OO (tãŋ$^{213}_{party}$; T)	IDENT-OO (tãŋ$^{21}_{party}$; T)
Weight	-1	-1	-1	-1
☞tãŋ^{21} -6	*		*****	
☞tãŋ^{213} -6		*****		*

Tableau 6.20. Paradigmatic leveling toward the full concave noun paradigm

3.3 Korean stem-final variations

The problem left in the Korean simulation based on PSI-OT-GLA is twofold: (1) The contextual variations in the experimental results, the preference of standard pronunciations in the locative/dative context in particular, could not be captured in PSI-OT-GLA, and (2) stem-final /t/→[tʃʰ] mappings were harmonically-bounded by /t/→[tʃ] mappings, so stem-final [tʃʰ] variants were generally impossible in PSI-OT-GLA. This brief section aims at a solution based on paradigmatic uniformity required in LIOO-HG.

In LIOO-HG, Korean stem-final variations can be thought as analogical extension driven by paradigmatic uniformity as proposed in Myers (2002), which requires a stem-final segment to be consistent between two suffixed forms if the correspondent bare forms also agree in the stem-final segment. For example, since the bare forms [ot] and [pat] agree in the stem-final segment [t], the suffixed forms should as well have the same stem-final segment. The stem-final [s] in the locative/dative form [os-e] could thus be extended to create the locative/dative form [pas-e] as a surface variation of the standard pronunciation [pat^h-e].[78] This correspondence relationship can be formalized with Myers's (2002) OO conjunction framework in (11) (see also §2).

(11) OO(ot, pat: $t/\_]_{Stem}$)^OO(os-e, pat^h-e: $s/\_]_{Stem}$): If the paradigms [ot] and [pat] agree in the segment [t] in a stem-final position, the paradigms [pat^h-e] and [os-e] must also agree in the segment [s] in a stem-final position.

Despite this paradigmatic uniformity pressure, the standard pronunciation of locative/dative forms in Korean remains dominant in the experiment data, suggesting a strong resistant force against the uniformity pressure. Recall that Korean locative/dative forms should have a higher token frequency since the locative/dative marker /-e/ is never omitted even in conversational speech. Such a high token frequency therefore strengthens the pressure to maintain the uniformity between an output and every token of a locative/dative paradigm like [pat^h-e] as required by the LIOO constraint in (12). The pressure of analogical extension is denoted by a higher weight of the OO conjuction constraint in Tableau 6.21, but the output candidate with the extension in fact has a lower raw score due to a great violation number of the LIOO constraint.

(12) IDENT-OO(pat^h-e; [cont]): The output and every token of [pat^h-e] must agree in the feature [cont].

/pat^h-e/ Bare: /pat/		OO(ot, pat: $t/\_]_{Stem}$)^ OO(os-e, pat^h-e: $s/\_]_{Stem}$)	IDENT-OO (pat^h-e; [cont])
Weight		-10	-1
☞pat^h-e	-10	*	
pas-e	-20		********************

Tableau 6.21. Preserving stem-final segments in locative/dative context in Korean

[78] In this analogical extension framework, it is also possible to level the stem-final segment of [os-e] to that of [pat^h-e] and create a surface variation [ot^h-e]. I will, however, leave a detailed discussion of directionality in LIOO-HG in future studies.

On the contrary, the accusative paradigm [pat^h-ɨl] has a significantly lower token frequency due to the trend of dropping the accusative marker; the uniformity pressure between the tokens of the accusative paradigm is insufficient to block the extension as demonstrated in Tableau 6.22. In sum, the blocking effect on the analogical extension in the locative/dative context can be accounted for with the strong uniformity pressure between the individual tokens of a locative/dative paradigm in the LIOO-HG model.

/pat^h-ɨl/ Bare: /pat/	OO(ot, pat: t/_]$_{Stem}$)^ OO(os-ɨl, pat^h-ɨl: s/_]$_{Stem}$)	IDENT-OO (pat^h-ɨl; [cont])
Weight	-10	-1
pat^h-ɨl -10	*	
☞pas-ɨl -5		*****

Tableau 6.22. Extended [t]~[s] alternation in accusative context

The solution to the problem caused by the extension of the [t]~[tʃʰ] alternation is straightforward with the introduction of OO conjunction constraint in LIOO-HG. For example, the source of a stem-final [tʃʰ] in the surface variant like [patʃʰ-ɨl] is the uniformity between [pat^h-ɨl] and [k'otʃʰ-ɨl] as required by the OO conjuction constraint in (13). With a higher constraint weight of the OO conjuction constraint and a low token frequency of [pat^h-ɨl] which cannot protect the identity of the accusative form, the stem-final segment in [pat^h-ɨl] is leveled to [tʃʰ] in [k'otʃʰ-ɨl] as an extension of the [t]~[tʃʰ] alternation Tableau 6.23. Since it is unnecessary to derive stem-final coronal obstruents from an underlying stem-final /t/ in LIOO-HG, the stem-final coronal obstruent types are not restricted by non-harmonically-bounded mappings.

(13) OO(k'ot, pat: t/_]$_{Stem}$)^OO(k'otʃʰ-ɨl, pat^h-ɨl: tʃʰ/_]$_{Stem}$): If the paradigms [k'ot] and [pat] agree in the segment [t] in a stem-final position, the paradigms [k'otʃʰ-ɨl] and [pat^h-ɨl] must also agree in the segment [tʃʰ] in a stem-final position.

/patʰ-ɨl/ Bare: /pat/	OO(k'ot, pat: t/_]$_{Stem}$)^ OO(k'otʃʰ-ɨl, patʰ-ɨl: tʃʰ/_]$_{Stem}$)	OO(patʰ-ɨl; [cont])
Weight	-10	-1
patʰ-ɨl -10	*	
☞patʃʰ-ɨl -5		*****

Tableau 6.23. Extended [t]~[tʃʰ] alternation in an accusative paradigm

3.4 Local summary

To sum up here, the absence of a complete lexical network is the primary cause of the drawbacks in PSI-OT-GLA. By expanding the lexical association in a rich lexicon with an OO constraint grammar, it is possible to offer thorough explanations for the acquisition and diachronic changes of morphophonology and improve the performance of simulating the production patterns generated by real speakers.

4. Additional advantages without input selection in LIOO-HG

This section aims to further explain why the OO approach should be preferred by discussing the possible issues encountered by PSI-OT-GLA when a non-surface-true UR seems absolutely necessary. The pair 'atom [ˈæɾə̃m]'~'atomic [əˈtʰɔm-ɪk]' in English is an ideal example of illustrating the need of such a non-surface-true UR: Since the two stem allomorphs ([ˈæɾə̃m] or [əˈtʰɔm]) cannot be derived from each other in any context, a non-surface-true UR [ætɔm] collecting the full vowel from both surface allomorphs seems necessary, and vowel reduction applies based on the stress position. As the segments of such non-surface-true URs usually forms a collection of the segments of their surface allomorphs, they will be referred to as 'superset UR' below.

A superset UR is impossible in PSI-OT-GLA since any non-surface-true input will have a zero token frequency and is thus an impossible input option. Therefore, the possible stem inputs for 'atom' are limited to [æɾə̃m] and [ətʰɔm], and PSI-OT-GLA predicts that if an output cannot surface from an input allomorph, surface variations emerge and diachronic changes occur. For instance, if /æɾə̃m+ɪk/ is selected as the input of 'atomic', the surface form cannot be [ətʰɔm-ɪk]. At least, there will no natural grammar to restore the schwa in the input specifically to [ɔ] or the flap to [tʰ] in the expected output. Some unusual surface variations of 'atomic' might thus be wrongly predicted by PSI-OT-GLA.[79]

Palauan (Flora 1974; Schane 1974) and Tonkawa (Hoijer 1933, 1949, Kissebirth 1970) are other languages that also require superset URs. In Palauan, full vowels coexist with a stress, and unstressed vowels are always reduced to a schwa as shown in (14).[80] As in the English case, it is impossible to derive all surface allomorphs by selecting any surface-true allomorph as the stem input since the full vowel in a stressed position is literally unpredictable. A superset UR that includes the full vowels in all surface allomorphs is thus required, which is /daŋob/ for 'cover opening' and /teʔib/ for 'pull out'.

(14) Unstressed vowel reduction in Palauan (Schane 1974:300)

Present Middle Verb	Future Participle (conservative)	Future Participle (innovative)	
[mə-dáŋəb]	[dəŋób-l]	[dəŋəb-áll]	'cover opening'
[mə-téʔəb]	[təʔíb-l]	[təʔəb-áll]	'pull out'

Syncope and apocope in Tonkawa also keep superset URs seemingly indispensable. As shown in the Tonkawa data in (15), stem vowels in different positions are deleted for different reasons: Apocope applies to avoid a word-final vowel, as in [notox_], and syncope applies to parse trochaic feet (H) of (HL) as in [(not_.xo).(n-oʔ)] and [(we-n_.to)(xo-ʔ)] (Gouskova 2003). Once again, a superset UR /notoxo/ is required for the stem morpheme as surface-true allomorphs cannot correctly derive all surface forms via vowel deletion and unpredictable vowel epenthesis.

(15) Syncope and apocope in Tonkawa

/notoxo/	[no.tox_]	'hoe'	cf. *[no.to.xo]
/notoxo+ʔ/	[not_.xo.-ʔ]	'he hoes it'	cf. *[no.to.xo-ʔ]
/we+notoxo+ʔ/	[we-n_.toxo-ʔ]	'he hoes them'	cf. *[we-no.to.xo-ʔ]

[79] One possible solution in PSI-OT-GLA is that a lexical input can be constructed through letter-to-sound associations; the spelling *atom* thus enables the input option similar to [ætɔm] to derive all surface allomorphs. See Tanenhaus et al. (1980) for general psycholinguistic evidence for a close association between orthography and sounds and Vendelin & Peperkemp (2006) for the influence in loanword adaption. This solution is nevertheless not possible for languages without a writing system, and is therefore not a general approach to deal with the problem.

[80] See also Crosswhite (2001, 2004) for a phonetic account of vowel reduction.

LIOO-HG does not have the above superset UR problem since the input of each paradigm is always equal to the target output, and the inputs in (15) are thus /notox/, /notxo-ʔ/, and /we-ntoxo-ʔ/ respectively. The goal of phonotactic learning is simply to acquire a phonological grammar that can produce inputs faithfully. This rich lexicon assumption does not contradict the acquisition of native phonotactics as mentioned in §4 of Ch. 1. In a framework that follows restrictive learning principles (e.g. initial Markedness » LIOO ranking bias), patterns that are absent in the learning inputs in Tonkawa, like word-final vowels and non-trochaic feet, will be forbidden by default. Furthermore, the lack of superset URs might be able to explain why there are exceptions to the Tonkawa syncope process found in Hoijer (1949) in (16): Learners may not be able to invent superset URs and thus unable to fully acquire the syncope patterns.

(16) Exceptions to Tonkawa syncope (see Appendix E for a full list)

/ʔaw+atak/	[ʔa.w-a.tak]	'deer'	cf. *[ʔaw.-tak]
/ʔawas+atak/	[ʔa.wa.s-a.tak]	'buffalo'	cf. *[ʔaw.s-a.tak]
/nekame+an/	[ne.ka.m-an]	'bone'	cf. *[nek.m-an]
/kala+kopul/	[ka.la.-ko.pul]	'round mouth'	cf. *[kal.-ko.pul]

Another challenge to PSI-OT-GLA is phonologically conditioned allomorph selection. That is, the choice of stem allomorphs is determined upon whether the combination between a stem allomorph and an affix violates any phonotactics. For example, the verb 'call' in Polish has two allomorphs /zEva/ and /zEv/ where /E/ stands for a yer alternation between /e/ or zero. As claimed in Rubach & Booji's (2001), when the verb is followed by an infinitive suffix [-tɕ] the input stem allomorph is /zEva/, which surfaces as [zva-tɕ] 'to call' with the suffix as in (17). If the suffix is third person plural [-õ], the input stem allomorph is /zEv/, and the suffixed form surface as [zv-õ] 'they call'.[81] Rubach & Booji propose that /zEva/ is not selected before [-õ] because the output *[zva.-õ] creates an onsetless syllable and thus violates higher-ranked ONSET. In Rubach & Booji's OT analysis, both verb allomorphs are listed at the input level, and their possible outputs can be evaluated simultaneously as in Tableau 6.24.

[81] According to Rubach & Booji, the vowel /e/ cannot be realized as the output of the yer alternation in both forms.

(17) Phonologically conditioned allomorph selection in Polish

/zEva+tɕ/	[zva-tɕ]	'to call'
/zEv+õ/	[zv-õ]	'they call'

/zEva/ +/-õ/ /zEv/	ONSET	COMPLEXONSET
☞zv-õ		*
zva.-õ	*!	

Tableau 6.24. Allomorph listing in phonologically conditioned allomorph selection

PSI-OT-GLA cannot account for this allomorph selection process since the input selection process is independent from the grammar evaluation process, and if the input of 'they call' is probabilistically selected as /zEva+õ/, the output might deviate from [zv-õ] depending on the ranking of MAX-IO, DEP-IO, and ONSET. When inputs are always identical to their outputs via Lexicon Optimization in LIOO-HG, the input of 'to call' and 'they call' is always identical to their output (i.e. /zva-tɕ/ and /zv-õ/), and input listing is redundant.

To conclude here, the OO model can incorporate the lexical parameters included in PSI-OT-GLA to have the same advantages as PSI-OT-GLA as shown in §2, but with an expanded lexical network it should not have the same flaws in PSI-OT-GLA in modeling Dutch, Mandarin, and Korean morphophonology (see §3). This section further illustrates how the OO model can avoid the potential challenges faced by PSI-OT-GLA if it is ultimately applied to the morphophonological acquisition of a wide variety of languages. Hopefully, after demonstrating that PSI-OT-GLA had a good start by assigning a more important role to the lexicon in previous chapters, the above brief discussion can bring up a more convincing conclusion that the OO model can further improve along the same line.

5. Closing remarks

This dissertation pursues a morphophonological acquisition model with a rich lexicon which, albeit being adopted in various usage-based theories on the basis of psycholinguistic evidence, has been marginalized in the mainstream of morphophonological studies in past decades. In particular, while learners traditionally are assumed to be innately equipped with some 'reasoning' ability to generalize abstract lexical representations, a morphophonological acquisition model based on this reasoning ability has been underdeveloped and perhaps generated very limited predictions.

This dissertation thus pleads for reconsideration for the lexical economy hypothesis and underscores the flexibility in morphophonology as revealed during morphophonological learning and changes and allowed with a rich lexical space. During morphophonological learning, the surface allomorphs are stored and referred to by learners to test grammar assumptions and produce surface variations. Diachronic developments can also be driven when learners gradually refer to different stored allomorphs over generations. These morphophonological patterns cannot be captured without releasing the power for a learning model to store and use abundant morphophonological information in a rich lexicon, which I assume to best approximate the real learning mechanism. Computational simulations were also implemented in this dissertation to compare the performance of a rich lexicon model and real speakers/learners, and the similar performance shown in this dissertation should sufficiently endorse the proposal of a morphophonology with a rich lexicon and following expansions in the same direction.

I believe that the languages and morphophonological patterns that have been addressed in this dissertation only comprise of a small proportion of those demanding an explanation based on lexical richness, and thus urge review and analysis of previous and existing phonological and morphophonological research questions from the perspective advocated in this dissertation. Along this line, the search for a more explanatory (morpho-)phonological learning mechanism and perhaps a desirably universal account of synchronic and diachronic (morpho-)phonological variation may not have to take unnecessary detours.

Bibliography

Albright, A. (2002). *The Identification of Bases in Morphological Paradigms*. PhD thesis, UCLA.

Albright, A. (2008). Explaining universal tendencies and language particulars in analogical change. In J. Good (Ed.), *Linguistic universals and language change* (pp. 144–181). Oxford, UK: Oxford University Press.

Albright, A. (2010). A Restricted Model of UR Discovery: Evidence from Lakhota. Ms., MIT.

Albright, A., & Kang, Y. (in press). Predicting innovative alternations in Korean verb paradigms. In *Proceedings of CIL18: The 18th International Congress of Linguists*. Amsterdam: John Benjamins.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, *90*(2), 119–161.

Albright, A., & Hayes, B. (2011). Learning and learnability in phonology. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (2nd ed., pp. 661–690). Oxford, UK: Blackwell.

Alderete, J., Tupper, P., & Frisch, S. A. (2013). Phonological constraint induction in a connectionist network: learning OCP-Place constraints from data. *Language Sciences*, *37*, 52–69.

Andersen, H. (1973). Abductive and deductive change. *Language*, *49*(4), 765–793.

Andersen, H. (1978). Perceptual and conceptual factors in abductive innovations. In J. Fisiak (Ed.), *Recent Developments in Historical Phonology* (pp. 1–22). The Hague: Mouton.

Anderson, J. R., & Ross, B. H. (1980). Evidence against a semantic-episodic distinction. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 441–466.

Angluin, D. (1978). Inductive inference of formal languages from positive data. *Information and Control*, *45*(2), 117–135.

Anttila, A. (1997). Deriving variation from grammar. In F. Hinskens, R. van Hout, & W. L. Wetzels (Eds.), *Variation, Change and Phonological Theory* (pp. 35–68). Amsterdam: John Benjamins.

Anttila, A. (2002). Morphologically conditioned phonological alternations. *Natural Language & Linguistic Theory*, *20*(1), 1–42.

Apoussidou, D. (2006). On-line learning of underlying forms. *ROA 835*. Retrieved from http://roa.rutgers.edu/article/view/845

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, *5*(2), 183–207.

Archangeli, D., & Douglas, P. (1994). *Grounded Phonology*. Cambridge, MA: The MIT Press.

Aksu-Koç, A. A., & Slobin, D. I. (1985). The acquisition of Turkish. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition, Vol. 1: the Data*. Hillsdale: Lawrence Erlbaum Associates.

Baayen, R. H., McQueen, J. M., & Dijkstra, T. (2003). Frequency effects in regular inflectional morphology: revisiting Dutch plurals. In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 355–390). Berlin: Mouton de Gruyter.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baer-Henney, D., & van de Vijver, R. (2012). On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology*, *3*(2), 221–249.

Baker, C. L. (1979). Syntactic Theory and the Projection Problem. *Linguistic Inquiry*, *10*(4), 533–581.

Bao, Z. (1999). *The structure of tone*. Oxford, UK: Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Bolker, B., Maechler, M., & Walker, S. (2013). lme4: Linear mixed-effect models using S4 classes. R package version 0.999999-2. Retrieved from http://cran.r- project.org/web/packages/lme4/index.html

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic Press.

Becker, M. (2009). *Phonological trends in the lexicon: the role of constraints*. PhD thesis, University of Massachusetts, Amherst.

Becker, M., & Tessier, A.-M. (2011). Trajectories of faithfulness in child-specific phonology. *Phonology*, *28*(02), 163–196.

Beckman, J. N. (1999). *Positional Faithfulness: An Optimality Theoretic Treatment of Phonological Asymmetries*. New York: Routledge.

Benua, L. (1995). Identity effects in morphological truncation. In J. N. Beckman, L.

W. Dickey, & S. Urbanczyk (Eds.) *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory*, 77–136. Amherst, MA: University of Massachusetts, Amherst.

Benua, L. (1997). *Transderivation identity: Phonological relations between words*. PhD thesis, University of Massachusetts, Amherst.

Berent, I. (2013). The phonological mind. *Trends in Cognitive Sciences*, *17*(7), 319–327.

Berent, I., Marcus, G. F., Shimron, J., & Gafos, A. I. (2012). The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition*, *83*(2), 113–139.

Berent, I., Wilson, C., Marcus, G., & Bemis, D. (2012). On the role of variables in phonology: Remarks on Hayes and Wilson (2008). *Linguistic Inquiry*, *43*(1), 97–119.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Berman, R. A. (1985). The acquisition of Hebrew. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition, Vol. 1: the Data* (pp. 255–371). Hillsdale, NJ: Lawrence Erlbaum Associates.

Boersma, P. (1998). *Functional Phonology*. PhD thesis, University of Amsterdam.

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & W. L. Wetzels (Eds.), *Loanword phonology* (pp. 11–58). Amsterdam: John Benjamins.

Boersma, P., & Hayes, B. (2001). Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, *32*(1), 45–86.

Boersma, P., & Pater, J. (2008). Convergence properties of a gradual learning algorithm for Harmonic Grammar. *ROA 970*. Retrieved from http://roa.rutgers.edu/article/view/1000

Boersma, P., & Weenick, D. (2013). Praat: doing phonetics by computer. Retrieved from http://www.fon.hum.uva.nl/praat/

Burzio, L. (2002). Missing players: Phonology and the past-tense debate. *Lingua*, *112*(3), 157–199.

Burzio, L. (2005). Sources of paradigm uniformity. In L. J. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory* (pp. 65–106). Oxford, UK: Oxford University Press.

Bybee, J. L. (1985). *Morphology: a study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford, CA: CSLI.

Bybee, J. L. (2001). *Phonology and Language Use*. Cambridge, UK: Cambridge University Press.

Bybee, J. L. (2002). Word frequency and context of use in lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Bybee, J. L., & Moder, C. L. (1983). Morphological Classes as Natural Categories. *Language*, *59*(2), 251–270.

Byun, T. M. (2011). A gestural account of a child-specific neutralisation in strong position. *Phonology*, *28*(3), 371–412.

Chao, Y.-R. (1930). A system of tone letters. *Le Maitre Phonetique*, *30*(1), 24–27.

Charles-Luce, J., & Luce, P. a. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*(1), 205–215.

Chen, J.-Y., & Chen, T.-M. (2002). Word-Form Encoding in Mandarin Chinese as Assessed by the Implicit Priming Task. *Journal of Memory and Language*, *46*(4), 751–781.

Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge, UK: Cambridge University Press.

Chen, T.-Y. (2008). Revisiting Yantai Tone Sandhi. *Toronto Working Paper in Linguistics 28: Proceedings of the ICEAL*, 1–14.

Chen, T.-Y. (2010). Some remarks on Contour Tone Units. *Journal of East Asian Linguistics*, *19*(2), 103–135.

Chen, T.-Y. (2013). Slope-based tonal phonology: a case study of Hakha Lai. Paper presented at *Chicago Linguistic Society 49*.

Chen, T.-Y., & Tucker, B. V. (2013). Sonorant onset pitch as a perceptual cue of lexical tones in Mandarin. *Phonetica*, *70*(3), 207–239.

Cho, S.-B. (1967). *A phonological study of Korean: With a historical analysis*. Uppsala, Sweden: Almqvist and Wiksells.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*(2), 193–228.

Choi, H. (2004). *A survey of standard pronunciation III* [phyocunpalum silthecosa III]. Seoul: The National Academy of Korean Language.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper

and Row.
Churma, D. G. (1982). Rule inversion in Chadic: a closer look. *Studies in African Linguistics*, *13*(1), 11–29.
Clahsen, H., & Neubauer, K. (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers. *Lingua*, *120*(11), 2627–2637.
Coetzee, A. W., & Kawahara, S. (2013). Frequency biases in phonological variation. *Natural Language & Linguistic Theory*, *31*(1), 47–89.
Colina, S. (2013). Galician geada: In defense of underspecification in Optimality Theory. *Lingua*, *133*(null), 84–100.
Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*(2), 274–284.
Crosswhite, K. M. (2001). *Vowel Reduction in Optimality Theory*. New York: Routledge.
Crosswhite, K. M. (2004). Vowel reduction. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based Phonology* (pp. 191–231). Cambridge, UK: Cambridge University Press.
Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121.
Da, J. (2004). Chinese text computing: syllable frequencies with tones. Retrieved from http://lingua.mtsu.edu/chinese-computing/
Davidson, L., Jusczyk, P., & Smolensky, P. (2004). The Initial and Final States: Theoretical Implications and Experimental Explorations of Richness of the Base. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in Phonological Acquisition* (pp. 321–368). Cambridge, UK: Cambridge University Press.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.
Do, Y. A. (2012). Learning Alternations in Korean Noun Paradigms. In E. Choi, A. Hogue, J. Punske, D. Tat, J. Schertz, & A. Trueman (Eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics* (pp. 319–327). Somerville, MA: Casadilla Press.
Duanmu, S. (1994). Against contour tone units. *Linguistic Inquiry*, *25*(4), 555–608.
Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed.). Oxford, UK: Oxford University Press.
Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels

in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578.

Eisenstat, S. (2009). *Learning Underlying Forms With MaxEnt*. Master thesis, University of Pennsylvania.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America*, *97*(3), 1893.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*(2), 419–439.

Flemming, E. (2003). The relationship between coronal place and vowel backness. *Phonology*, *20*(3), 335–373.

Flora, M. (1974). *Palauan phonology and morphology*. PhD thesis, UCSD.

Fon, J., Chiang, W.-Y., & Cheung, H. (2004). Production and perception of the two dipping tones (Tone 2 and Tone 3) in Taiwan Mandarin. *Journal of Chinese Linguistics*, *32*(2), 249–280.

Fowler, A. E. (1990). How early phonological development might set the stage for phoneme awareness. In S. A. Brady & D. Shankweiler (Eds.), *Phonological Processes in Literacy* (pp. 97–117). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fry, D. (1966). The development of the phonological system in the normal and deaf child. In F. Smith & G. A. Miller (Eds.), *The genesis of language* (pp. 187–206). Cambridge, MA: The MIT Press.

Fu, Q.-J., & Zeng, F.-G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech Language and Hearing*, *5*(1), 45–57.

Gahl, S. (2008). "Time" and "Thyme" Are Not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Language*, *84*(3), 474–496.

Garding, E., Kratochvil, P., Svantesson, J.-O., & Zhang, J. (1986). Tone 4 and Tone 3 Discrimination in Modern Standard Chinese. *Language and Speech*, *29*(3), 281–293.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology. Learning, memory, and cognition*, *22*(5), 1166–83.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *17*(1), 152–162.

Gouskova, M. (2003). *Deriving Economy: Syncope in Optimality Theory*. PhD thesis, University of Massachusetts, Amherst.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 501–518.

Greenberg, J. (1966). *Language universals, with special reference to feature hierarchies* (Janua Linguarum, Series Minor 59). The Hague: Mouton.

Greenlee, M., & Ohala, J. J. (1980). Phonetically motivated parallels between child phonology and historical sound change. *Language Sciences*, *2*(2), 283–308.

Guy, G. R. (2007). Lexical Exceptions in Variable Phonology. *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 109–120.

Guy, G. R. (2014). Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. *Lingua*, *142*, 57–65.

Hale, K. (1973). Deep-surface canonical disparities in relation to analysis and change: An Australian example. In T. Sebeok (Ed.), *Current Trends in Linguistics 11* (pp. 401–458). The Hague: Mouton.

Hale, M., & Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, *29*(4), 656–683.

Hall, T. A., & Hamann, S. (2006). Towards a typology of stop assibilation. *Linguistics*, *44*(6), 1195–1236.

Hall, T. A., Hamann, S., & Zygis, M. (2006). The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association*, *36*(1), 59.

Halle, M. (1985). Speculations about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 101–114). New York: Academic Press.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly Progress Report* (Research Laboratory of Electronics, MIT), *101*, 198–212.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences*, *17*(3), 111–120.

Hare, M. L., Ford, M., & Marslen-Wilson, W. D. (2001). Ambiguity and frequency effect in regular verb inflection. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 181–200). Amsterdam: John Benjamins.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal*

of Linguistics, *42*(1), 25–70.
Hayes, B. (2000). Phonological Restructuring in Yidiŋ and its Theoretical Consequences. In B. Hermans & M. van Oosterdorp (Eds.), *The Derivational Residue in Phonological Optimality Theory* (pp. 175–206). Amsterdam: John Benjamins.
Hayes, B. (2004). Phonological acquisition in Optimality Theory: the early stages. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in Phonological Acquisition* (pp. 158–203). Cambridge, UK: Cambridge University Press.
Hayes, B., Steriade, D., & Kirchner, R. (2004). *Phonetically Based Phonology*. Cambridge, UK: Cambridge University Press.
Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry*, *44*(1), 45–75.
Hayes, B., & Wilson, C. (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry*, *39*(3), 379–440.
Hockett, C. F. (1947). Peiping Phonology. *Journal of the American Oriental Society*, *67*(4), 253–267.
Hoijer, H. (1933). *Tonkawa. An Indian language of Texas*. New York: J. J. Augustin.
Hoijer, H. (1949). *An Analytical Dictionary of the Tonkawa Language. University of California Publications in Linguistics, Vol. 5, No. 1*. Berkeley, CA: University of California Press.
Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (Ed.), *Current progress in historical linguistics* (pp. 95–105). Amsterdam: North Holland.
Hooper, J. B. (1980). Child morphology and morphophonemic change. In J. Fisiak (Ed.), *Historical Morphology* (pp. 157–188). The Hague: Mouton.
Horwood, G. (2006). Association Faith and Korean Palatalization. In E. Baković, J. Ito, & J. J. McCarthy (Eds.), *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince* (pp. 111–138). Santa Cruz, CA: Linguistic Research Center, Department of Linguistics, University of California at Santa Cruz.
Huang, T. (2001). The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. *OSU Working Papers in Linguistics*, *55*, 23–42.
Hyman, L. M. (1970). How Concrete Is Phonology ? *Language*, *46*(1), 58–76.
Hyman, L. M. (1975). Review of A Synchronic Phonology of Mandarin Chinese by

Chin-chuan Cheng. *Journal of Chinese Linguistics*, *3*(1), 88–99.

Hyman, L. M. (2011). Does Gokana really have no syllables? Or: what's so great about being universal? *Phonology*, *28*(01), 55–85.

Inkelas, S. (1994). The consequences of Optimization for Underspecification. *ROA 40*.

Inkelas, S., & Rose, Y. (2007). Positional Neutralization: A Case Study from Child Language. *Language*, *83*(4), 707–736.

Ito, Junko & Armin Mester. (1995). Japanese phonology. In John Goldsmith (ed.), *The Handbook of Phonological Theory 1st Edition*. Oxford: Blackwell. Pp. 817–38.

Jakobson, R. (1929). *Remarques sur l'évolution phonologique du russe comparée avec celle des autres langues slaves (=Travaux du Cercle Linguistique de Prague 2)*. Prague: Jednota C‡eskoslovensk´ych Matematiku (a Fysiku(.

Jarosz, G. (2006a). Richness of the Base and Probabilistic Unsupervised Learning in Optimality Theory. In *Proceedings of the Eighth Meeting of the ACL Special Interest Group on Computational Phonology* (pp. 50–59).

Jarosz, G. (2006b). *Rich lexicons and restrictive grammars - Maximum Likelihood Learning in Optimality Theory*. PhD thesis, Johns Hopkins Univeristy.

Jarosz, G. (2007). Restrictiveness and Phonological Grammar and Lexicon Learning. *Proceedings from the Annual Meeting of the Chicago Linguistic Society 43 vol. 1,* 125–139.

Jarosz, G. (2011). The Roles of Phonotactics and Frequency in the Learning of Alternations. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development* (pp. 321–333). Somerville, MA: Casadilla Press.

Jarosz, G. (to appear). Naive Parameter Learning for Optimality Theory-the Hidden Structure Problem. In *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*. Amherst, MA: GLSA, University of Massachusetts, Amherst.

Jesney, K., & Tessier, A.-M. (2011). Biases in Harmonic Grammar: the road to restrictive learning. *Natural Language & Linguistic Theory*, *29*(1), 251–290.

Johnson, K. (1997). Speaker perception without speaker normalization. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 1–8). San Diego, CA: Academic Press.

Joseph, L. (1997). *Handbook of Palauan Grammar (Vol. 1)*. Koror: Palau Ministry

of Education.
Joseph, L. (1999). *Handbook of Palauan Grammar (Vol. 2)*. Koror: Palau Ministry of Education.
Jun, E. (2002). An experimental study of the effect of release of English syllable final stops on vowel epenthesis in English loanwords. In *Studies in Phonetics, Phonology and Morphology 8* (Vol. 8, pp. 117–134). Seoul: Hankwukmunhwasa.
Jun, J. (2010). Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics*, *19*(2), 137–179.
Jun, J., & Lee, J. (2007). Multiple stem-final variants in Korean native nouns and loanwords. *Eoneohag*, *47*, 159–187.
Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. PhD thesis, Ohio State University.
Jusczyk, P. W., Cutler, a, & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child development*, *64*(3), 675–87.
Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants ́ Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402–420.
Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, *43*(3), 253–291.
Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics*, *2*(2), 161–180.
Kang, E., Lee, H.-Y., & Kim, J. (2004). The phonetic realization of syllable codas in Korean. *Journal of the Korean Society of Phonetic Science and Speech Technology*, *49*, 1–30.
Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology*, *20*(2), 219–273.
Kaplan, A. (2010). *Phonology shaped by phonetics: The case of intervocalic lenition*. PhD thesis, University of California, Santa Cruz.
Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319.
Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Kenstowicz, M., & Kisseberth, C. W. (1977). *Topics in phonological theory*. New York: Academic Press.

Kenstowicz, M., & Kisseberth, C. W. (1979). *Generative phonology: description and theory*. New York: Academic Press.

Kerkhoff, A. (2007). *Acquisition of morpho-phonology: The Dutch voicing alternation*. PhD thesis, University of Utrecht.

Kim, H. (2001). A phonetically based account of phonological stop assibilation. *Phonology*, *18*(1), 81–108.

Kim, H., & Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, *24*(3), 295–312.

Kim, H.-G., Niimi, S., & Hirose, H. (1993). Devoicing of Vowel in Korean. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, *27*, 151–155.

Kim, M.-R., & Duanmu, S. (2004). "Tense" and "Lax" Stops in Korean. *Journal of East Asian Linguistics*, *13*(1), 59–104.

Kim, S. (2003). *A survey of standard pronunciation II* [phyocunpalum silthecosa II]. Seoul: The National Academy of Korean Language.

Kiparsky, P. (1973). Phonological representations. In O. Fujimura (Ed.), *Three Dimensions of Linguistic Theory* (pp. 3–136). Tokyo: TEC.

Kiparsky, P. (1982a). *Explanation in phonology*. Dordrecht: Foris.

Kiparsky, P. (1982b). Lexical Phonology and Morphology. In Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm: Selected Papers from SICOL-1981* (pp. 3–91). Seoul: Hanshin Publishing Company.

Kiparsky, P. (1985). Some consequences of Lexical Phonology. *Phonology Yearbook*, *2*, 85–138.

Kirchner, R. (1999). Preliminary thoughts on "phonologisation" within an exemplar-based speech processing system. *UCLA Working Papers in Linguistics: Papers in Phonology 2*, *1*, 205–231.

Kirchner, R. (2001). *An Effort-Based Approach to Consonant Lenition*. New York: Routledge.

Kirchner, R. (2011). Modelling exemplar-based phonologization. In A. C. Cohn, C. Fourgeron, & M. K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 332–346). Oxford, UK: Oxford University Press.

Kirchner, R., Moore, R. K., & Chen, T.-Y. (2010). Computing phonological generalization over real speech exemplars. *Journal of Phonetics*, *38*(4), 540–

547.

Kisseberth, C. W. (1970). Vowel elision in Tonkawa and derivational constraints. In J. M. Sadock & A. L. Vanek (Eds.), *Studies presented to Robert B. Lees by his students* (pp. 109–137). Edmonton, AB and Champaign, IL: Linguistic Research.

Kochetov, A. (2011). Palatalization. In M. Van Oostendorp, C. J. Ewen, E. V. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology Volume III: Phonological Processes* (pp. 1666–1690). Oxford, UK: Blackwell.

Krämer, M. (2012). *Underlying Representation*. Cambridge, UK: Cambridge University Press.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, *66*(6), 1668.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*(2-3), 263–285.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 338–362). Berlin: Mouton de Gruyter.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59.

Leben, W. R. (1973). *Suprasegmental phonology*. PhD thesis, MIT.

Lee, C.-Y. (2007). Does Horse Activate Mother? Processing Lexical Tone in Form Priming. *Language and Speech*, *50*(1), 101–123.

Lee, I. (1999). *A principles-and-parameters approach to the acquisition of (the morphosyntax of) IP in Korean*. PhD thesis, University of Essex.

Legendre, G., Miyata, Y., & Smolensky, P. (1990). *Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: An application*. Technical report #90-4, Institute of Cognitive Science, University of Colorado at Boulder.

Legendre, G., Miyata, Y., & Smolensky, P. (2006). *Harmonic Grammar - A Formal Multi-Level Connectionist Theory of Linguistic Well-Formedness: Theoretical Foundations*. Technical report #90-5, Institute of Cognitive Science, University of Colorado at Boulder.

Levelt, C. C., Schiller, N. O., & Levelt, W. J. M. (1999). A Developmental Grammar for Syllable Structure in the Production of Child Language and Languages.

Brain and Language, *68*(1-2), 291–299.

Lin, H.-B. (1988). *Contextual Stability of Taiwanese Tones*. PhD thesis, University of Connecticut.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin Lexical Tones when F0 Information is Neutralized. *Language and Speech*, *47*(2), 109–138.

Lubowicz, A. (2002). Derived environment effects in Optimality Theory. *Lingua*, *112*(4), 243–280.

Lombardi, L. (2001). Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In L. Lombardi (Ed.), *Segmental phonology in Optimality Theory: Constraints and representation* (pp. 13–45). Cambridge, UK: Cambridge University Press.

Lowry, R. (2001-2013). Significance of the Difference Between Two Correlation Coefficients. Retrieved from http://www.vassarstats.net/rdiff.html

Matisoff, J. A. (1988). *The dictionary of Lahu*. Berkeley, CA: University of California Press.

Macken, M. A. (1980). The Child's Lexical Representation: The "Puzzle-Puddle-Pickle" Evidence. *Journal of Linguistics*, *16*(1), 1–17.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, *43*(1/2), 1–123.

MacWhinney, B. (1991/2000). *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magri, G. (2012). Convergence of error-driven ranking algorithms. *Phonology*, *29*(2), 213–269.

Magri, G., & Storme, B. (2013). A closer look at Boersma and Hayes (2001) simulations: extensions, analyses, implications. Paper presented at *Chicago Linguistic Society 49*.

Mańczak, W. (1980). Laws of analogy. In J. Fisiak (Ed.), *Historical Morphology* (pp. 283–288). The Hague: Mouton.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). *Overregularization in Language Acquisition. Monographs of the Society for Research in Child Development 57.*

Martin, S. E. (1992). *A reference grammar of Korean*. Seoul: Hanshin Publishing Company.

Martinez-Gil, F. (1997). Word-final epenthesis in Galician. In F. Martinez-Gil & A. Morales-Front (Eds.), *Issues in the Phonology and Morphology of the Major*

Iberian Languages (pp. 269–340). Washington, D.C.: Georgetown University Press.

McCarthy, J. J. (2005). Optimal paradigm. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory* (pp. 170–210). Oxford, UK: Oxford University Press.

McCarthy, J. J., & Prince, A. (1995). Faithfulness and reduplicative identity. *ROA 60*. Retrieved from http://roa.rutgers.edu/article/view/61.

McEnery, T., & Xiao, R. (2003-2008). The Lancaster Corpus of Mandarin Chinese. Retrieved from http://www.lancaster.ac.uk/fass/projects/corpus/LCMC/

Mei, T. (1977). Tones and tone sandhi in 16th century Mandarin. *Journal of Chinese Linguistics*, *5*(2), 237–260.

Menn, L. (1980). Child phonology and phonological theory. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology: Vol. 1, Production* (pp. 23–42). New York: Academic Press.

Merchant, N., & Tesar, B. (2005). Learning underlying forms by searching restricted lexical subspaces. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, *41*(2), 1–15.

Mikheev, A. (1997). Automatic rule induction for unknown-word guessing. *Computational Linguistics*, *23*(3), 405–423.

Moreton, E., & Pater, J. (2012a). Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass*, *6*(11), 686–701.

Moreton, E., & Pater, J. (2012b). Structure and Substance in Artificial-Phonology Learning, Part II: Substance. *Language and Linguistics Compass*, *6*(11), 702–718.

Myers, J. (2002). Analogical Modeling: An Exemplar-Based Approach to Language. In R. Skousen, D. Londsdale, & D. B. Parkinson (Eds.), *Analogical Modeling: An exemplar-based approach to language* (pp. 265–300). Amsterdam: John Benjamins.

Myers, J., & J. Tsay (2013). Modeling universal and lexical influences on phonotactic judgments. Paper presented at The 4th International Theoretical Phonology Conference, National Chengchi University, Taipei.

Neu, H. (1980). Ranking of constraints on –t,d deletion in American English. In W. Labov (Ed.), *Locating language in time and space* (pp. 37–54). New York: Academic Press.

Nosofsky, R. M. (1988). On Exemplar-Based Exemplar Representations: Reply to

Ennis (1988). *Journal of Experimental Psychology: General*, *117*(4), 412–414.

Ohala, J. J. (1975). Phonetic explanation for nasal sound patterns. In C. A. Ferguson, L. M. Hyman, & J. J. Ohala (Eds.), *Nasálfest* (pp. 289–316). Stanford, CA: Stanford University Press.

Ohala, J. J. (1978). Southern Bantu vs. the World: the case of palatalization of labials. *Berkeley Linguistic Society 4*, 370–386.

Ohala, J. J., & Lorentz, J. (1977). The Story of [w]: An Exercise in the Phonetic Explanation for Sound Patterns. *Proceedings from the Annual Meeting of the Berkeley Linguistic Society*, *3*, 577–599.

Papen, R. (1998). French: Canadian varieties. In J. Edwards (Ed.), *Language in Canada* (pp. 160–176). Cambridge, UK: Cambridge University Press.

Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition*, *6*(3), 201–253.

Pater, J. (2000). Nonuniformity in English stress: the role of ranked and lexically specific constraints. *Phonology*, *17*(2), 237–274.

Pater, J. (2008). Gradual learning and convergence. *Linguistic Inquiry*, *39*(2), 334–346.

Pater, J. (2009). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In S. Parker (Ed.), *Phonological argumentation: Essays on evidence and motivation*. London, UK: Equinox.

Pater, J., Staubs, R., Jesney, K., & Smith, B. (2012). Learning probabilities over underlying representations. In *12th Meeting of the Special Interest Group on Computational Morphology and Phonology 2012 (SIGMORPHON 2012)* (pp. 62–71). New York: Curran Associates, Inc.

Peng, S. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics*, *25*(3), 371–400.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress "deafness." In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.

Peperkamp, S., & Dupoux, E. (2007). Learning the mapping from surface to underlying representations in an artificial language. In J. Cole & I. J. Haulde (Eds.), *Laboratory Phonology 9* (pp. 315–338). Berlin: Mouton de Gruyter.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–158). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101–140). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*(4), 516–530.

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, *25*(2), 143–170.

Port, R. (2010). Rich memory and distributed phonology. *Language Sciences*, *32*(1), 43–55.

Prince, A. (2002). Entailed Ranking arguments. *ROA 500*. Retrieved from http://roa.rutgers.edu/article/view/510

Prince, A. (2007). The pursuit of theory. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 33–60). Cambridge, UK: Cambridge University Press.

Prince, A., & Smolensky, P. (2004). *Optimality Theory: constraint interaction in generative grammar*. Oxford, UK: Blackwell. [Originally published in 1993]

R Team. (2012). R: a language and environment for statistical computing. Version 2.14.2. Retrieved from http://www.r-project.org/

Riggle, J. (2004). *Generation, Recognition, and Learning in Finite State Optimality Theory*. PhD thesis, UCLA.

Rubach, J., & Booij, G. E. (2001). Allomorphy in Optimality Theory: Polish Iotation. *Language*, *77*(1), 26–60.

Sapir, E. (1933). La réalité psychologique des phonèmes. *Journal de psychologie normale et pathologique*, *30*, 247–265.

Schane, S. A. (1974). How abstract is abstract? *Papers from parasession on Natural Phonology, Proceedings from the Annual Meeting of the Chicago Linguistic Society*, 10, 297–317.

Scharinger, M., & Idsardi, W. J. (2014). Sparseness of vowel category structure: Evidence from English dialect comparison. *Lingua,140*, 35–51.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, *3*, 83–109.

Shuai, L., Li, B., & Gong, T. (2012). Priming effects of tones and segments in lexical processing in mandarin. Paper presented at *the 6th International Conference on Speech Prosody*.

Silva, D. J. (1992). *The phonetics and phonology of stop lenition in Korean*. PhD thesis, Cornell University.

Simões, M. C. P., & Stoel-Gammon, C. (1979). The acquisition of inflections in Portuguese: a study of the development of person markers on verbs. *Journal of Child Language*, *6*(01), 53–67.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental science*, *12*(6), 914–9.

Smolensky, P. (1996). *The Initial State and 'Richness of the Base 'in Optimality Theory*. Ms., Johns Hopkins University.

Stampe, D. (1979). *A Dissertation on Natural Phonology*. New York: Garland Publishing.

Steriade, D. (1999). Lexical conservatism in French adjectival liaison. In J.-M. Authier, B. E. Bullock, & L. A. Reed (Eds.), *Formal Perspectives in Romance Linguistics,* (pp. 243–270). Amsterdam: John Benjamins.

Steriade, D. (2001a). Lexical conservatism and the notion base of affixation. Ms., UCLA.

Steriade, D. (2001b). The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. Ms., UCLA.

Stoel-Gammon, C. (1998). The role of babbling and phonology in early linguistic development. In A. M. Wetherby, S. F. Warran, & M. E. Fey (Eds.), *Communication and language intervention series vol. 7: transitions in prelinguistic communication* (pp. 87–110). Baltimore: Paul H. Brookes.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of child language*, *38*(1), 1–34.

Swingley, D., & Aslin, R. N. (2002). Lexical Neighborhoods and the Word-Form Representations of 14-Month-Olds. *Psychological Science*, *13*(5), 480–484.

Swingley, D. (2003). Phonetic Detail in the Developing Lexicon. *Language and Speech*, *46*(2-3), 265–294.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition*, *8*(6), 513–520.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *2*(3), 339–363.

Telfer, C. (2006). *Coronalization as Assibilation*. Master thesis, University of Calgary.

Tesar, B. (1995). *Computational Optimality Theory*. PhD thesis, University of Colorado, Boulder.

Tesar, B. (2004). Using Inconsistency Detection to Overcome Structural Ambiguity. *Linguistic Inquiry*, *35*(2), 219–253.

Tesar, B. (2008). Learning Phonological Grammars for Output-Driven Maps. *ROA 1043*.

Tesar, B., Alderete, J., Horwood, G., Merchant, N., Nishitani, K., & Prince, A. (2003). Surgery in Language Learning. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics 22* (pp. 477–490). Somerville, MA: Casadilla Press.

Tesar, B., & Prince, A. (2003). Using phonotactics to learn phonological alternations. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, *39*(2), 241–269.

Tessier, A.-M. (2007). *Biases and stages in OT phonological acquisition*. PhD thesis, University of Massachusetts, Amherst.

Tessier, A.-M. (2009). Frequency of violation and constraint-based phonological learning. *Lingua*, *119*(1), 6–38.

The Eclipse Foundation. (2014). Eclipse. Version 4.3.2. Retrieved from http://eclipse.org/

Trubetzkoy, N. S. (1929). *Polabische Studien. Sitzungsberichte der Akademie der Wissenschaften in Wien, Philosophisch-historische Klasse, Bund 211, Abhandlung 4*. Wien/Leipzig: Hölder-Pichler-Tempsky A.-G.

Trubetzkoy, N. S. (1936). Essai d'une théorie des oppositions phonologiques. *Journal de psychologie normale et pathologique*, *33*, 5–18.

van de Weijer, J. (2012). *Grammar as Selection: Combining Optimality Theory and Exemplar Theory*. Nagoya: Kougaku Shuppan.

van de Weijer, J. (2014). The origin of OT constraints. *Lingua*, *142*, 66–75.

van Oostendorp, M. (2014). Selective lexicon optimization. *Lingua*, *142*, 76–84.

Vendelin, I., & Peperkamp, S. (2006). The influence of orthography on loanword adaptations. *Lingua*, *116*(7), 996–1007.

Vennemann, T. (1972). Rule inversion. *Lingua*, *29*, 209–242.

Vennemann, T. (1974). Restructuring. *Lingua*, *33*(1), 137–156.

Walley, A. C. (1993). The Role of Vocabulary Development in Children's Spoken Word Recognition and Segmentation Ability. *Developmental Review*, *13*(3),

286–350.
Waterson, N. (1971). Child phonology: a prosodic view. *Journal of Linguistics*, *7*(2), 179–211.
Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*(3), 247–274.
Wedel, A. B. (2007). Feedback and regularity in the lexicon. *Phonology*, *24*(01), 147–185.
Weerman, F. (1993). The diachronic consequences of first and second language acquisition: the change from OV to VO. *Linguistics*, *31*(5), 903–932.
Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.
Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition*. Cambridge, MA: The MIT Press.
Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*(1), 25–47.
Yang, C. (2003). *Knowledge and learning in natural language*. Oxford, UK: Oxford University Press.
Yang, C. (2005). On productivity. *Linguistic Variation Yearbook*, *5*(1), 265–302.
Yip, M. (1980). *The tonal phonology of Chinese*. PhD thesis, MIT.
Yip, M. (1996). Lexicon Optimization in Languages without Alternations. In J. Durand & B. Laks (Eds.), *Current Trend in Phonology: Models and Methods Vol.2* (pp. 757–788). Manchester: ESRI, University of Salford Publication.
Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.
Yun, G., & Jackson, S. R. (2004). The Phonetic Signs of Categorical Variation in Korean Stop Phonology. *Harvard Studies in Korean Linguistics*, *10*, 303–316.
Zamuner, T. S., Gerken, L., & Hammond, M. (2005). The acquisition of phonology based on input: a closer look at the relation of cross-linguistic and child language data. *Lingua*, *115*(10), 1403–1426.
Zamuner, T. S., Kerkhoff, A., & Fikkert, P. (2006). Acquisition of voicing neutralization and alternations in Dutch. In *BUCLD 30* (pp. 701–712). Somerville, MA: Casadilla Press.
Zamuner, T. S., Kerkhoff, A., & Fikkert, P. (2012). Phonotactics and morphophonology in early child language: Evidence from Dutch. *Applied Psycholinguistics*, *33*(3), 481–499.

Zhang, J. (2002). *The effects of duration and sonority on contour tone distribution*. New York: Routledge.

Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In B. Hayes, D. Steriade, & R. Kirchner (Eds.), *Phonetically based Phonology* (pp. 157–190). Cambridge, UK: Cambridge University Press.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*(1), 153–201.

Zhang, J., Lai, Y., & Craig, S. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua*, *121*(2), 181–206.

Zhang, J., Lai, Y., & Turnbull-Sailor, C. (2006). Wug-testing the "tone circle" in Taiwanese. *WCCFL 25*, 453–461.

Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, *68*(3), 161–191.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.

Zoll, C. (2002). Optimal Tone Mapping. *Linguistic Inquiry*, *34*(2), 225–268.

Appendix A. Index of Constraints

Faithfulness constraints:

DEP-IO: Every output segment must have an input correspondent. (Ch. 2:§5.2)

DEP-OO(*P*): The segments in the output must have a correspondent in every token of a lexical paradigm *P*. (Ch. 6:§3.1)

FAITH-IO_{ROOT}: Every phonological element in the input of a root morpheme must be preserved in the output. (Ch. 2:§2.3)

FAITH-IO_{AFFIX}: Every phonological element in the input of an affix morpheme must be preserved in the output. (Ch. 2:§2.3)

IDENT(σ_1)-IO: Segmental features in the first syllable must be preserved in the output. (Ch. 5:§1.2)

IDENT(asp)-IO: Input [asp] specification must be preserved in the output. (Ch. 5:§1.1)

IDENT(cont)-IO: Input [continuant] specification must be preserved in the output. (Ch. 5:§1.2)

IDENT(del rel)-IO: Input [delayed release] specification must be preserved in the output. (Ch. 5:§1.2)

IDENT(strid)-IO: Input [strid] specification must be preserved in the output. (Ch. 5:§1.1)

IDENT(tense)-IO: Input [tense] specification must be preserved in the output. (Ch. 5:§1.1)

IDENT(voi)-IO: Every input specification of [voice] must be preserved in the output. (Ch. 2:§2.1)

IDENT-OO(*P*;F): The output and each paradigm *P* of a lexical category must agree in the features F. (Ch. 6:§2.2)

IDENT-OO(pat^h-e;[cont]): The output and every token of [pat^h-e] must agree in the feature [cont]. (Ch. 6:§3.3)

IDENT-OO($tãŋ^{21}_{block}$;T): The output and every token of [$tãŋ^{21}$] 'block' must agree in the tonal specification. (Ch. 6:§3.2)

IDENT-OO($tãŋ^{21}_{party}$;T): The output and every token of [$tãŋ^{21}$] 'political party' must agree in the tonal specification. (Ch. 6:§3.2)

IDENT-OO($tãŋ^{213}_{block}$;T): The output and every token of [$tãŋ^{213}$] 'block' must agree in the tonal specification. (Ch. 6:§3.2)

IDENT-OO($tãŋ^{213}_{party}$;T): The output and every token of [$tãŋ^{213}$] 'political party' must agree in the tonal specification. (Ch. 6:§3.2)

MAX-σ́: Every input segment of a stressed syllable must have an output correspondent. (Ch. 2:§2.4)

MAX-IO: Every input segment must have an output correspondent. (Ch. 2:§5.2)

MAX-LINK: An input autosegment must have an output correspondent. (Ch. 4:§1)

MAX-LINK(H/ML_): An input H autosegment preceded by autosegments ML must have an output correspondent (i.e. *MLH→ML). (Ch. 4:§1)

MAX-LINK(L/M_H): An input L autosegment between autosegments M and H must have an output correspondent (i.e. *MLH→MH). (Ch. 4:§1)

MAX-LINK(M): An input M autosegment must have an output correspondent (i.e. *MLH→H and *MLH→HL). (Ch. 4:§1)

MAX-OO(*P*): The segments in every token of a lexical paradigm *P* must have a correspondent in the output. (Ch. 6:§3.1)

OO(*a*,*c*:F)^OO(*b*,*d*:G): If the forms *a* and *c* agree in a set of features F, the forms *b* and *d* must also agree in a set of feature G. (Ch. 6:§2)

OO(k'ot, pat: t/_]$_{Stem}$)^OO(k'otʃ^{h}-ɨl, pat^{h}-ɨl: tʃ^{h}/_]$_{Stem}$): If the paradigms [k'ot] and [pat] agree in the segment [t] in a stem-final position, the paradigms [k'otʃ^{h}-ɨl] and [pat^{h}-ɨl] must also agree in the segment [tʃ^{h}] in a stem-final position. (Ch. 6:§3.3)

OO(ot, pat: t/_]$_{Stem}$)^OO(os-e, pat^{h}-e: s/_]$_{Stem}$): If the paradigms [ot] and [pat] agree in the segment [t] in a stem-final position, the paradigms [os-e] and [pat^{h}-e] must also agree in the segment [s] in a stem-final position. (Ch. 6:§3.3)

OO([+animal];F): The output and each paradigm with the semantic feature [+animal] must agree in the features F. (Ch. 6:§2.5)

OO('dog (pl.)';F): The output and each paradigm of 'dog (pl.)' must agree in the features F. (Ch. 6:§2.5)

OO(root='dog';F): The output and each paradigm with the root morpheme 'dog' must agree in the features F. (Ch. 6:§2.5)

Markedness constraints:

*/*P*/→*M*: A paradigm *P* as the input of a morpheme *M* is prohibited. (Ch. 2:§6.2)

*[+cont, -strid][-son]: A cluster of a non-strident fricative followed by a non-sonorant is prohibited. (Ch. 2:§2.1)

*+asp: Aspirated consonants are prohibited. (Ch. 5:§1.1)

*+strid: Stridents are prohibited. (Ch. 5:§1.1)

*+tense: Tense consonants are prohibited. (Ch. 5:§1.1)

*+asp/_]σ: Syllable-final aspirated consonants are prohibited. (Ch. 5:§1.1)

*+strid/_]σ: Syllable-final stridents are prohibited. (Ch. 5:§1.1)

*+tense/_]σ: Syllable-final tense consonants are prohibited. (Ch. 5:§1.1)

[Dor]/_]$_{\text{STEM}}$=[k]: Stem-final labial stops must be [k](i.e. plain velar stop). (Ch. 5:§3.2)

[Lab]/_]$_{\text{STEM}}$=[p]: Stem-final labial stops must be [p] (i.e. plain bilabial stop). (Ch. 5:§3.2)

*CODA: Syllable codas are prohibited. (Ch. 2:§5.2)

FINALCONCAVE: In a phrase-final position, a tone must surface as a concave tone. (Ch. 4:§4.2)

LAZY: Affricate (****) > Strident Fricative (***) >Stop (**) >Non-strident Fricative (*). (Ch. 5:§1.2)

*LONGLAPSE: MLML (or four adjacent [-high] autosegments) sequence is prohibited in the output. (Ch. 4:§1)

MOVE-AS-UNIT: (Ch. 2:§2.1)

*MLH: MLH is prohibited in the output. (Ch. 4:§1)

*NONFINAL-MLH: MLH in a non-phrase-final position is prohibited in the output. (Ch. 4:§1)

NONFINALDIPPING: In a non-final position, a tone must surface as ML. (Ch. 4:§4.2)

OCP-MLH: Two adjacent MLH contour tone units are prohibited. (Ch. 4:§1)

ONSET: Syllables without onset are prohibited. (Ch. 2:§5.2)

*PEAK/SEG: A segment category SEG as syllable peak is prohibited. (Ch. 2:§2.3)

*t/_]$_{\text{STEM}}$-S: A stem-final /t/ is prohibited when preceding a suffix. (Ch. 5:§3.1)

*Ti: A sequence of coronal stop + high front vowel is prohibited. (Ch. 5:§1.2)

*Tɨ: A sequence of coronal stop + high central vowel is prohibited. (Ch. 5:§1.2)

*V[-voi]V: Intervocalic voiceless segments are prohibited. (Ch. 2:§2.1)

*V[+voi]V: Intervocalic voiced segments are prohibited. (Ch. 2:§2.1)

*VOICEDOBS: Voiced obstruents are prohibited. (Ch. 2:§5.2)

*VOICEDOBSCODA: Voiced obstruent codas are prohibited. (Ch. 2:§5.2)

*VtV: Intervocalic /t/s are prohibited. (Ch. 2:§2.1)

Appendix B. ERCs of each possible input in the Dutch simulations

Harmonically-bounded output candidates will not be included in the following ERCs.

Input: /bɛd/ Target: [bɛt]	ONSET	*VOIOBS	*VOIOBSCODA	*CODA	*V[-voi]V	MAX-IO	DEP-IO	IDENT(voi)-IO
bɛt~bɛd	e	W	W	e	e	e	e	L
bɛt~bɛ	e	e	e	L	e	W	e	L
bɛt~ɛ	W	L	e	L	e	W	e	L
bɛt~pɛt	e	L	e	e	e	e	e	W
bɛt~pɛ	e	L	e	L	e	W	e	W
bɛt~bɛ.ti	e	W	e	L	e	e	W	L
bɛt~pɛ.ti	e	L	e	L	W	e	W	W
bɛt~pɛ.di	e	e	e	L	e	e	W	W

Input: /bɛt/ Target: [bɛt]	ONSET	*VOIOBS	*VOIOBSCODA	*CODA	*V[-voi]V	MAX-IO	DEP-IO	IDENT(voi)-IO
bɛt~bɛ	e	e	e	L	e	W	e	e
bɛt~ɛ	W	L	e	L	e	W	e	e
bɛt~pɛt	e	L	e	e	e	e	e	W
bɛt~pɛ	e	L	e	L	e	W	e	W
bɛt~bɛ.di	e	W	e	L	e	e	W	W
bɛt~bɛ.ti	e	e	e	L	W	e	W	e
bɛt~pɛ.ti	e	L	e	L	W	e	W	W
bɛt~pɛ.di	e	e	e	L	e	e	W	W

Input: /bɛd+ən/ Target: [bɛ.d-ən]	ONSET	*VOIOBS	*VOIOBSCODA	*CODA	*V[-voi]V	MAX-IO	DEP-IO	IDENT(voi)-IO
bɛ.d-ən~pɛ.d-ən	e	L	e	e	e	e	e	W
bɛ.d-ən~pɛ.t-ən	e	L	e	e	W	e	e	W
bɛ.d-ən~ɛ.-ən	W	L	e	e	e	W	e	e
bɛ.d-ən~bɛ.d-ə	e	e	e	L	e	W	e	e
bɛ.d-ən~pɛ.t-ə	e	L	e	L	W	W	e	W
bɛ.d-ən~pɛ.d-ə	e	L	e	L	e	W	e	W
bɛ.d-ən~ɛ.-ə	W	L	e	L	e	W	e	e
bɛ.d-ən~bɛ.d-ə.ni	e	e	e	L	e	e	W	e
bɛ.d-ən~pɛ.t-ə.ni	e	L	e	L	W	e	W	W
bɛ.d-ən~pɛ.d-ə.ni	e	L	e	L	e	e	W	W
bɛ.d-ən~ɛ.-ə.ni	W	L	e	L	e	W	W	e

Input: /bɛt+ən/ Target: [bɛ.d-ən]	ONSET	*VOIOBS	*VOIOBSCODA	*CODA	*V[-voi]V	MAX-IO	DEP-IO	IDENT(voi)-IO
bɛ.d-ən~bɛ.t-ən	e	L	e	e	W	e	e	L
bɛ.d-ən~pɛ.d-ən	e	L	e	e	e	e	e	W
bɛ.d-ən~pɛ.t-ən	e	L	e	e	W	e	e	e
bɛ.d-ən~ɛ.t-ən	W	L	e	e	W	W	e	L
bɛ.d-ən~ɛ.-ən	W	L	e	e	e	W	e	L
bɛ.d-ən~bɛ.t-ə	e	L	e	L	W	W	e	L
bɛ.d-ən~bɛ.d-ə	e	e	e	L	e	W	e	e
bɛ.d-ən~pɛ.t-ə	e	L	e	L	W	W	e	e
bɛ.d-ən~pɛ.d-ə	e	L	e	L	e	W	e	W
bɛ.d-ən~ɛ.t-ə	W	L	e	L	W	W	e	L
bɛ.d-ən~ɛ.-ə	W	L	e	L	e	W	e	e
bɛ.d-ən~bɛ.t-ə.ni	e	e	e	L	W	e	W	L
bɛ.d-ən~bɛ.d-ə.ni	e	e	e	L	e	e	W	e
bɛ.d-ən~pɛ.t-ə.ni	e	L	e	L	W	e	W	W
bɛ.d-ən~pɛ.d-ə.ni	e	L	e	L	e	e	W	W
bɛ.d-ən~ɛ.t-ə.ni	W	L	e	L	W	W	W	L
bɛ.d-ən~ɛ.-ə.ni	W	L	e	L	e	W	W	e

Input: /ɑp/ Target: [ɑp]	ONSET	*VOIOBS	*VOIOBSCODA	*CODA	*V[-voi]V	MAX-IO	DEP-IO	IDENT(voi)-IO
ɑp~tɑp	L	e	e	e	e	e	W	e
ɑp~tɑ	L	e	e	L	e	W	e	e
ɑp~ɑ	e	e	e	L	e	W	e	e
ɑp~tɑ.pi	L	e	e	L	W	e	W	e
ɑp~tɑ.pi	L	W	e	L	e	e	W	W
ɑp~ɑ.pi	e	e	e	L	W	e	W	e
ɑp~ɑ.bi	e	W	e	L	e	e	W	W

Input: /ɑp+ən/ Target: [ɑ.p-ən]	ONSET	*VOIOBS	*VOIOBSCODA	*CODA	*V[-voi]V	MAX-IO	DEP-IO	IDENT(voi)-IO
ɑ.p-ən~tɑ.p-ən	L	e	e	e	e	e	W	e
ɑ.p-ən~ɑ.p-ə	e	e	e	L	e	W	e	e
ɑ.p-ən~tɑ.p-ə	L	e	e	L	e	W	W	e
ɑ.p-ən~tɑ.b-ən	L	W	e	e	L	e	W	W
ɑ.p-ən~ɑ.b-ə	L	W	e	L	L	W	e	W
ɑ.p-ən~tɑ.b-ə	L	e	e	L	L	W	W	W
ɑ.p-ən~tɑ.p-ə.ni	L	e	e	L	e	e	W	e
ɑ.p-ən~ɑ.p-ə.ni	e	e	e	L	e	e	W	e
ɑ.p-ən~tɑ.b-ə.ni	L	W	e	L	L	e	W	W
ɑ.p-ən~ɑ.b-ə.ni	e	e	e	L	L	e	W	W

Appendix C. Tone 1, Tone 2, and Tone 4 fillers in the elicitation task

Tone 1		Tone 2		Tone 4	
汤 [tʰaŋ^{55}]	‘soup’	鞋 [ɕie^{24}]	‘shoe’	烫 [tʰãŋ^{51}]	‘hot’
飞 [fei^{55}]	‘fly’	连 [liɛ̃n^{24}]	‘link’	弱 [ɻuɔ^{51}]	‘weak’
亲 [tɕʰĩn^{55}]	‘kiss’	牛 [niou^{24}]	‘cow’	毕 [pi^{51}]	‘complete’
东 [tɔ̃ŋ^{55}]	‘east’	熟 [ʂou^{24}]	‘ripe’	跳 [tʰiau^{51}]	‘jump’
听 [tʰĩŋ^{55}]	‘listen’	茄 [tɕʰiɛ^{24}]	‘eggplant’	去 [tɕʰy^{51}]	‘go’
失 [ʂɨ^{55}]	‘lose’	提 [tʰi^{24}]	‘lift’	让 [ɻãŋ^{51}]	‘let’
车 [tʂʰɤ^{55}]	‘car’	肥 [fei^{24}]	‘fat’	妹 [mei^{51}]	‘sister’
操 [tsʰau^{55}]	‘grasp’	农 [nɔ̃ŋ^{24}]	‘agriculture’	踏 [tʰa^{51}]	‘step on’
科 [kʰɤ ^{55}]	‘subject’	旁 [pʰãŋ^{24}]	‘side’	进 [tɕĩn^{51}]	‘forward’
他 [tʰa^{55}]	‘he’	紅 [hɔ̃ŋ^{24}]	‘red’	谢 [ɕiɛ^{51}]	‘thank’
衣 [i^{55}]	‘cloth’	田 [tʰiɛ̃n^{24}]	‘farm’	掛 [kua^{51}]	‘hang on’
灾 [tsai^{55}]	‘disaster’	俗 [su^{24}]	‘convention’	厚 [hou^{51}]	‘thick’
凶 [ɕiɔ̃ŋ^{55}]	‘evil’	佛 [fou^{24}]	‘Buddha’	莫 [muɔ^{51}]	‘do not’
精 [tɕĩŋ^{55}]	‘precise’	强 [tɕʰiãŋ^{24}]	‘strong’	现 [ɕiɛ̃n^{51}]	‘current’

Appendix D. Production distributions of individual stems in Generation 5 of the simulated results and Jun & Lee's experimental results

Numbers represent proportions (%) of each stem-final variation for each stem. Exp = Jun & Lee's experiment results, Sim = simulation results.

'sickle' /nas/	nominative *-i*					accusative *-ɨl*					locative/dative *-e*				
	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	100	0	0	0	0	100	0	0	0	0	100	0	0	0	0
Sim	97	0	0	3	0	95.9	0	1.4	2.7	0	96.7	0	1	2.3	0

'clothes' /os/	nominative *-i*					accusative *-ɨl*					locative/dative *-e*				
	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	100	0	0	0	0	90	10	0	0	0	100	0	0	0	0
Sim	82.5	0	0	17.3	0	80.5	0	6	13.5	0	76.2	0	8.5	15.3	0

'day' /natʃ/[82]	nominative *-i*					accusative *-ɨl*					locative/dative *-e*				
	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	10	0	0	90	0	40	0	0	60	0	0	0	0	100	0
Sim	36.2	1.5	0	61.4	0	34.8	0.9	6.6	56.9	0	34.3	0	7.6	56.8	0

'breast' /tʃətʃ/[83]	nominative *-i*					accusative *-ɨl*					locative/dative *-e*				
	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	50	0	0	50	0	90	0	0	10	0	None				
Sim	31.4	0	0	68.5	0	30.2	0.1	3.8	65.6	0					

'flower' /k'otʃʰ/[84]	nominative *-i*					accusative *-ɨl*					locative/dative *-e*				
	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	20	80	0	0	0	40	60	0	0	0	0	90	10	0	0
Sim	34.1	65.6	0	0	0	25.8	68.2	5.9	0	0	25.2	66.8	7.8	0	0

[82] There are some output tokens with a rare onset change as a natural consequence of probabilistic constraint rankings, which compose of 0.9% in the nominative context, 0.5% in the accusative context, and 1.1% in the locative/dative context.

[83] Output tokens with a rare onset change occupy 0.1%, 0.3%, and 0.2% in the three contexts.

[84] Output tokens with an onset change are 0.3%, 0.1%, and 0.2% in the three contexts.

‘face’	nominative *-i*					accusative *-il*					locative/dative *-e*				
/natʃʰ/	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	30	70	0	0	0	40	60	0	0	0	45.5	54.5	0	0	0
Sim	**0.1**	**99.9**	**0**	**0**	**0**	**0.1**	**99.6**	**0.3**	**0**	**0**	**0.4**	**99.6**	**0**	**0**	**0**

‘field’	nominative *-i*					accusative *-il*					locative/dative *-e*				
/patʰ/	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	10	90	0	0	0	20	80	0	0	0	0	10	90	0	0
Sim	**52.3**	**27.1**	**0**	**20.6**	**0**	**46.7**	**16.7**	**18.8**	**17.8**	**0**	**33.2**	**2.7**	**47.7**	**16.4**	**0**

‘red bean’	nominative *-i*					accusative *-il*					locative/dative *-e*				
/pʰatʰ/	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t	s	tʃʰ	tʰ	tʃ	t
Exp	45.5	54.5	0	0	0	60	30	10	0	0	45.5	9.1	36.4	0	9.1
Sim	**51.3**	**0**	**0**	**48.7**	**0**	**43.6**	**0**	**18.3**	**37.5**	**0**	**36.6**	**0**	**23.9**	**39.5**	**0**

‘kitchen’	nominative *-i*			accusative *-il*			locative/dative *-e*		
/puəkʰ/	plain	asp	tense	plain	asp	tense	plain	asp	tense
Exp	80	20	0	90	10	0	60	40	0
Sim	**100**	**0**	**0**	**100**	**0**	**0**	**100**	**0**	**0**

‘outside’	nominative *-i*			accusative *-il*			locative/dative *-e*		
/pak’/	plain	asp	tense	plain	asp	tense	plain	asp	tense
Exp	0	0	100	0	0	100	0	10	90
Sim	**100**	**0**	**0**	**100**	**0**	**0**	**100**	**0**	**0**

‘wall’	nominative *-i*			accusative *-il*			locative/dative *-e*		
/pjək/	plain	asp	tense	plain	asp	tense	plain	asp	tense
Exp	100	0	0	100	0	0	100	0	0
Sim	**100**	**0**	**0**	**100**	**0**	**0**	**100**	**0**	**0**

‘leaf’	nominative *-i*			accusative *-il*			locative/dative *-e*		
/ipʰ/	plain	asp	tense	plain	asp	tense	plain	asp	tense
Exp	0	100	0	30	70	0	0	100	0
Sim	**100**	**0**	**0**	**100**	**0**	**0**	**100**	**0**	**0**

‘rice’ /pap/	nominative *-i*			accusative *-il*			locative/dative *-e*		
	plain	asp	tense	plain	asp	tense	plain	asp	tense
Exp	100	0	0	100	0	0	100	0	0
Sim	**100**	**0**	**0**	**100**	**0**	**0**	**100**	**0**	**0**

Appendix E. Possible exceptions to Tonkawa syncope from Hoijer (1949)

The survey of exceptions below is limited to monomorphemic or suffixed forms in Hoijer (1949).

Initial (LL) allowed without syncope in morphologically complex words

ʔa.w-a.tak	‘deer’	(*ʔaw.-tak)	No. 8.1
ʔa.wa.s-a.tak	‘buffalo’	(*ʔaw.s-a.tak)	No. 10.1
ʔa.so.y-ey.la.pan	‘elm tree’	(*ʔas.y-ey.la.pan)	No. 16
ta.xa.c-e:.kin	‘all that day’	(*tax.c-e:.kin)	No. 154.4
ne.ka.m-an	‘bone’	(*nek.m-an)	No. 282.1
ne.xa.l-an	‘a snore’	(*nex.l-an)	No. 292.1
ka.la.-ya.mas	‘lips’	(*kal.-yamas)	No. 362.1
ka.la.-ko.pul	‘round mouth’	(*kal.-ko.pul)	No. 362.2
Cf. kal.-ʔok	‘mouth hair’	(*ka.la.-ʔok)	No. 362.3
wa.wa.n-an	‘throat’	(*waw.nan)	No.421.1
Cf. yat.k-an	‘frozen man’	(*ya.ti.kan)	No. 462.1
ya.can.-an	‘heart’	(*yac.n-an)	No. 468.1
Cf. yan.t-an	‘wind’	(*ya.na.t-an)	No. 469
ha.ni.l-es.ʔow	‘mouse’	(*han.les.ʔow)	No. 571.1
ha.ko.x-an	‘tired’	(*hak.xan)	No. 576.1
ha.ko.c-an	‘smoke’	(*hak.can)	No. 578.1
he.ʔe.-ca	‘that place’	(*heʔ.-ca)	No. 622.1
he.ma.y-an	‘ghost’	(*hem.yan)	No. 629.1

Initial (LL) or (LH) allowed in monomorphemic words

ʔa.wa.hey	‘Pawnee Indians’	No. 11
ʔa.sa:.hey	‘left’	No. 14
ʔa.so:.ka	‘sugar’	No. 17
ʔa.le:.na	‘wheat, flour’	No. 23
pe.ne.tix.kaʔ	‘Comanche Indians’	No. 95
me.li.kan	‘American’ (Loanword?)	No. 124
ne.sa.wo.nan	‘horse’	No. 298
ka.na.ʔa.kay	‘the other side’	No. 349
wa.ʔa.nes	‘as soon as’	No. 419
wa.ʔa.say	‘one side of’	No. 420

ya.si.la.way	'lizard'	No. 514
he.ma.xan	'chicken'	No. 630
ho.ko.pak.xon	'hat'	No. 729
ho.xo.lo:.ko	'shell'	No. 731
ho.sa.ʔas	'(several) new'	No. 732