

# MODELLING THE ACQUISITION AND EVOLUTION OF VOWEL HARMONY

By

Frédéric Mailhot

A thesis submitted to  
the Faculty of Graduate Studies and Research  
in partial fulfillment of  
the requirements for the degree of

DOCTOR OF PHILOSOPHY

Institute of Cognitive Science

CARLETON UNIVERSITY

Ottawa, Ontario

April, 2010

© Copyright by Frédéric Mailhot, 2010

## **ABSTRACT**

This thesis is a computational investigation of the mechanisms by which vowel harmony may be acquired by learners in a synchronic setting, and evolve diachronically from a pre-harmonic stage. I present a pair of computational models, grounded in recent research in categorisation, speech perception, and computational language learning, that give explicit accounts of both of these phenomena.

The acquisition model, LIBPHON, demonstrates the viability of an instance-based approach to productivity and generalisation in phonological acquisition, and gives support to an emerging view of phonological knowledge as a set of generalisations over non-segmented acoustic lexical representations. The evolutionary model supports the broadly Ohalian notion that sound change is initiated at the individual level by weakly-biased listeners, and augmented via iterated cycles of transmission and acquisition.

Finally, this approach to phonological acquisition and change, viewed in a Marrian context, is argued to tie phonology more closely to adjacent fields in the cognitive sciences, and in fact to provide a paradigmatic example of how such interdisciplinary, multi-level research can be carried out.

## ACKNOWLEDGEMENTS

I have rewritten this section multiple times now, and still cannot find the words to adequately acknowledge the contributions that others have made to this thesis, whether material or intellectual. Most of what's right here is a result of someone else's suggestion, and most of what's wrong is a result of my failure to take other peoples' suggestions into account.

I have known for as long as I can remember that I wanted to go to grad school (although I certainly faltered many times along the path). That particular drive came from my parents, who taught me the value of education, who encouraged me to pursue my dreams, and who never stopped believing that I could pull this off (even when my progress seemed to indicate otherwise). In addition to this moral/emotional groundwork, they have unhesitatingly supported me financially, and in particular didn't bat an eye when I suggested that the 2007 LSA Summer Linguistic Institute might be a good place for me to cement a thesis topic. Without that particular trip, this would be a completely different (or, more likely, nonexistent) document, and my thinking on certain linguistic matters would still be stuck in the 20<sup>th</sup> century. I love them and thank them from the bottom of my heart.

Academically, my path has been rather haphazard, and retracing the steps here would only be boring.<sup>1</sup> In roughly chronological order (but almost cer-

---

<sup>1</sup>Ask me if we ever meet at a conference!

tainly subject to some unintended omissions), the following people have made comments or suggestions that either furthered my general development as a linguist, or the development of this thesis in particular: Anna Mushynski, Nigel Duffield, Brendan Gillon, John Lewis, Doina Precup, Ray Jackendoff<sup>2</sup>, Charles Reiss, Mark Hale, Dana Isac, Ash Asudeh, Ida Toivonen, Lev Blumenfeld, Jeff Mielke, Andrea Gormley, Alan Hogue, Andy Wedel, audiences at a variety of universities, and the attendees of the Ottawa-Carleton Phonology (OCP... get it?) reading group.

I was fortunate to be at Carleton for the last five years. The interdisciplinarity of a Cognitive Science department gave me the freedom to explore directions that might not have been available to me in a more traditional linguistics setting, and the general collegiality of the department made it an excellent setting. I'd like to particularly thank Clara and Andrea (my fellow linguists) for good discussion good times, David for asking questions that made my head hurt, Neal and Crystal for being good friends and good office-mates, and Edgar for being possibly the nicest person ever. Lianne, Colleen and May kept the administrative things to sides humming along smoothly, and were always quick to help and forgiving of my frequent forgettings. My committee, Ash, Lev and Jeff, made the experience of creating this thesis about as enjoyable as could be hoped for. Ash was a fantastic advisor, who knew just when you give me a kick in the pants, and when

---

<sup>2</sup>OK, we've never really met, but his 2002 book *Foundations of Grammar* rekindled my love of linguistics at a time when it was waning.

to ask the right question. Lev has an positively encyclopedic knowledge of matters phonological and helped keep me (somewhat!) grounded in matters of interest to practicing phonologists. Jeff's thesis and his seminar on features cemented some of my ways of thinking about language, and more generally he was always incredibly helpful with suggestions on both the form and substance of whatever I was trying to say. Many of the conference presentations I gave would have been hopeless without his intervention.

It was Anna, my then-friend-now-wife, who knew me well enough to suggest that I try a linguistics course, after I had abandoned my physics aspirations. Without those few words so long ago, or your unconditional love and support over the last eight years, this would have been an impossibility. Thank you for everything, this thesis is dedicated to you.

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Explaining phonological patterns . . . . .	1
1.2	Methods . . . . .	6
1.3	Outline . . . . .	9
<b>2</b>	<b>Background</b>	<b>11</b>
2.1	Linguistics . . . . .	11
2.1.1	Vowel harmony . . . . .	11
2.1.2	Neutral vowels . . . . .	12
2.1.3	Criteria for vowel harmony . . . . .	15
2.1.4	Coarticulation . . . . .	16
2.1.5	Perceptual compensation . . . . .	19
2.1.6	Phonologisation . . . . .	20
2.1.7	The origins of phonological assimilation . . . . .	21
2.1.8	On language and language change . . . . .	24
2.1.9	The actuation problem . . . . .	27

2.2	Instance-based approaches to cognition and perception . . .	28
2.2.1	Exemplar models in psychology . . . . .	29
2.2.2	Exemplar models in linguistics . . . . .	31
2.3	Computational and mathematical Background . . . . .	38
2.3.1	Types of models . . . . .	38
2.3.2	Information flow . . . . .	40
2.3.3	Instance-based models and lazy learning . . . . .	41
2.4	Previous/related work . . . . .	44
2.4.1	Formal accounts of vowel harmony . . . . .	44
2.4.2	Computational accounts of vowel harmony . . . . .	50
2.4.3	Language change modelling . . . . .	53
<b>3</b>	<b>Acquisition</b>	<b>59</b>
3.1	LIBPHON . . . . .	61
3.1.1	<i>Decisions &amp; mechanisms</i> redux . . . . .	62
3.2	The task . . . . .	70
3.2.1	The languages . . . . .	70
3.2.2	Perception/categorisation . . . . .	78
3.2.3	Production/generation . . . . .	79
3.3	Simulations . . . . .	83
3.3.1	Information flow . . . . .	83
3.4	Evaluation . . . . .	87
3.4.1	Generalisation . . . . .	87

3.4.2	Evaluative measures . . . . .	88
3.5	Discussion . . . . .	95
<b>4</b>	<b>Evolution</b>	<b>97</b>
4.1	Modelling the emergence of harmony . . . . .	98
4.1.1	The agent . . . . .	98
4.1.2	Simulations . . . . .	102
4.1.3	Biased toward success? . . . . .	105
4.2	Conclusions . . . . .	109
4.2.1	Coarticulation, <i>prise deux</i> . . . . .	111
<b>5</b>	<b>Discussion</b>	<b>113</b>
5.1	Phonology in cognitive science . . . . .	113
5.2	Phonology as cognitive science . . . . .	116
5.2.1	The Marrian approach . . . . .	116
5.2.2	The continued necessity of “traditional” phonology .	121
5.2.3	Categoricity and gradience in phonology . . . . .	123
5.3	Diachronic phonology as predictive science . . . . .	124
5.4	Learning . . . . .	125
<b>6</b>	<b>Conclusions and Future Directions</b>	<b>128</b>
6.1	Contributions . . . . .	129
6.2	Remaining work . . . . .	131
6.2.1	Acquisition and use . . . . .	131



6.2.2	Change . . . . .	133
6.2.3	LIB <sub>PHON</sub> . . . . .	134
<b>Appendix A Equations for formant synthesis</b>		<b>137</b>
<b>Appendix B Beta distribution</b>		<b>139</b>
<b>Appendix C <i>k</i>-means clustering</b>		<b>141</b>

## LIST OF TABLES

2.1	Finnish backness harmony . . . . .	12
2.2	Tangale tongue root harmony . . . . .	13
2.3	Wolof tongue root harmony . . . . .	14
3.1	Opaque neutrality in LIBPHON's languages . . . . .	78
3.2	Transparent neutrality in LIBPHON's languages . . . . .	78
3.3	Contents of 10-word lexicon . . . . .	94

## LIST OF FIGURES

3.1	Schematic view of linguistic agent . . . . .	62
3.2	Example of lexical entry for GIDEGEBI NOM . . . . .	73
3.3	Singular forms of GIDEGEBI . . . . .	75
3.4	Plural (opaque) forms of GIDEGEBI . . . . .	76
3.5	Plural (transparent) forms of GIDEGEBI . . . . .	77
3.6	Vertical flow . . . . .	84
3.7	Oblique flow . . . . .	85
3.8	Vertical flow with peers . . . . .	86
3.9	Oblique flow with peers . . . . .	86
3.10	RMSE <i>v.</i> lex size. Vertical, transparent. . . . .	90
3.11	RMSE <i>v.</i> lex size. Oblique, opaque. . . . .	91
3.12	GUBOGOBU: 10-word lexicon . . . . .	92
3.13	GUBOGOBU: 30-word lexicon . . . . .	93
4.1	Architecture of a linguistic agent . . . . .	98
4.2	Effects of varying coarticulation . . . . .	104

4.3	Vowel distribution in $F1 - F2$ space . . . . .	107
B.1	Beta distribution . . . . .	140

# LIST OF ALGORITHMS

1	PERCEIVE( <i>input</i> , <i>k</i> ) . . . . .	79
2	PRODUCE(LABEL, <i>k</i> ) . . . . .	81
3	Iterated learning of lexical harmony . . . . .	102

"[...] we would naturally turn to the study of uniformities in the output (formal and substantive universals), which we must then attribute to the structure of the [language acquisition] device (*or, if this can be shown, to uniformities in the input, this alternative rarely being a serious one* in the cases of interest)." (Chomsky, 1986)

---

"[...] the linguistic input is actually a far better source than a rich Universal Grammar for all the information which is necessary for the child to figure out alternations or distributions which innate grammatical principles of markedness would supposedly help her to discover." (Samuels, 2009)

---

# CHAPTER 1

## INTRODUCTION

### 1.1 EXPLAINING PHONOLOGICAL PATTERNS

In explaining the existence of typologically frequent synchronic sound patterns, generative phonologists typically suggest that humans come to the task of language acquisition equipped with a rich base of innate, domain-specific knowledge (Chomsky and Halle, 1968; Kiparsky, 2008). That is, the distribution of phonological patterns—both common and rare—across the world’s language is explained by (or at least constrained by) humans’ common, genetically endowed, *language-specific*, initial state. Our rich innate knowledge in turn leads to the development of (representationally and/or procedurally) rich synchronic grammars (*e.g.* the account of velar palatalisation in Halle et al. (2000), who appeal to an innate feature

geometry.)

However, several phonologists and phoneticians have noted that many—perhaps most—recurrent patterns can be given explanations grounded in articulatory and acoustic factors (Ohala, 1989; Hale and Reiss, 2000; Hayes et al., 2004). Researchers who adopt this functional, or “grounded” approach (Archangeli and Pulleyblank, 1994), divide according to whether they take the functional pressures to have synchronic or diachronic influence, that is, whether speakers adapt their outputs to maximize articulatory ease (Kirchner, 1998) and minimize risk of listener error (Lindblom, 1990), or whether these biases are more latent and ateleological, manifesting only through the multiplicative effect of successive iterations of transmission and acquisition. In fact, this kind of historical phonological explanation has a long history, dating back at least to the work of Baudouin de Courtenay (1895), who explicitly suggested that misperception of form due to persistent physical biases in production and perception could result in sound change and the emergence of regular synchronic patterns. A recent extended treatment of this type of diachronic explanation is given in Blevins’s (2004) *Evolutionary Phonology*. Blevins argues from basic principles of parsimony that the duplication of explanation (between synchronic and diachronic accounts) should be eliminated, and in particular that diachronic arguments take precedence:

“Principled diachronic explanations have priority over com-



peting synchronic explanations unless independent evidence demonstrates, beyond reasonable doubt, that a synchronic account is warranted.” (Blevins (2004),p.23)

The basis for claiming that diachrony has priority over synchrony in phonological explanation is the fact that there are multiple theoretically and empirically well-motivated diachronic channels through which linguistic regularities can emerge. Blevins (2004, pp.120–121) draws explicit parallels to biological evolution in motivating five sources of crosslinguistic similarity/typological regularity:

1. *Direct inheritance* Features inherited from a shared ancestral language (e.g. *r*-loss in British and Australian English)
2. *Natural selection* Independent “natural” (viz. phonetically-motivated) development (analogous to parallel evolution), e.g. of final obstruent devoicing in Turkic, Indo-European
3. *Physical and psychological constraints on form/function* Innate aspects of perception/production (e.g. categorical perception), and phonological universals
4. *“Non-natural” factors* Social or external forces (e.g. language contact, strong prescriptive norms, L2 learning)
5. *Chance* Random non-naturally motivated development of similar features (e.g. restriction to purely nasal consonants word-finally in

Japanese and Gilbertese)

This thesis aims to provide support for this mode of phonological explanation by investigating circumstances under which a relatively common and well-studied phonological pattern, *vowel harmony*, can arise, both synchronically within individuals, and diachronically in a population of speakers. My goals are threefold:

1. Provide support for a particular account of the diachronic emergence of vowel harmony in a population of speakers
2. Given the logical priority of acquisition in the diachronic explanatory scenario laid out above, provide support for a particular view of the acquisition and use of phonology
3. Show that this leads to a tighter integration between linguistics and adjacent fields in the cognitive sciences

In short, I will give evidence for the view of vowel harmony most commonly associated with John Ohala (Ohala, 1993), namely that the emergence and typological distribution of vowel harmony is straightforwardly accounted for by a diachronic process involving biased (noisy) transmission and acquisition by means of domain-general operations. My motivations for choosing this particular empirical domain stem from several factors:

1. On the first hand, vowel harmony serves as an ideal test-bed for theo-

ries of acquisition in phonology as its extant theoretical descriptions comprise nearly all facets of generative phonological theory, such as supra- and sub-segmental structure, spreading, targets, triggers, lexical generalisations, alternations.

2. Vowel harmony serves here as a particular exemplar of a more general phenomenon: phonetically natural phonological assimilation. The success of the approach here will directly motivate its extension to all such phenomena (*e.g.* patterns of nasal assimilation or voicing assimilation in consonant clusters).
3. Goldsmith (2009) introduces the term *kernels*<sup>1</sup> to refer to language-specific (*i.e.* atheoretical) phonological generalisations accumulated over years of study, and suggests that they are the pieces of knowledge against which new phonological theories are assessed.

In phonology, these *kernels* [...] are what we know about things like sound inventories, vowel harmony, syllables and feet, stress systems, tone systems, and so on: the things by which we test and inspect any new theory. (Goldsmith, 2009, p.3)

Goldsmith goes on to suggest that the connectionist phonology introduced in the early 1990s (*cf.* Hare, 1990; Gasser and Lee, 1992;

---

<sup>1</sup>Which rather unfortunately overlaps non-trivially with a machine-learning method closely related to the instance-based approach I develop here.

Wheeler and Touretzky, 1993) failed to gain any traction exactly because it failed to engage with the traditional kernels of phonological knowledge. Although empiricist accounts of vowel harmony exist (*cf.* Hare, 1990; Hayes and Londe, 2006; Goldsmith and Riggle, 2007, among others), there are (to the best of my knowledge) no existing implemented exemplar-based models of the acquisition and use of vowel harmony. This thesis, then, is an initial foray into the kernels of phonological knowledge against which the model it proposes must be assessed.

## 1.2 METHODS

The key methodological tool put to use in the research presented here is *computational modelling*. In approaching any scientific problem, there may be several motivations for choosing to develop a computational model:

1. The problem at hand is too difficult to solve analytically (or no analytic description of the problem exists).
2. An approximate theoretical result might not be reliable, and it is necessary to check it with a different method.
3. An experiment is expensive or not feasible to perform.

In the present case, each of these is relevant. We shall see below that there are analytic models of language change, but their domains of application differ in important ways from the problem I am examining, which has no analytic description, as far as I know. In addition, the claims that I am investigating are about changes that occur over centuries and millenia. Although it would be in principle possible to test a prediction empirically by waiting for a sufficiently long time, in practice we would like some indication of whether the predictions are borne out on a shorter time scale.

More mundanely, modelling as methodology gives the researcher a “virtual lab” that allows for much tighter control over parameters of interest than experimentation with human subjects, which in turn leads to the possibility of perfect repeatability/replication. Also, computational models generate quantitative data, which at least in principle allows for comparison and (ideally) choice between competing accounts.

Note, finally, that models with particular parametric specifications are almost trivially falsifiable; if a given parameter setting fails to generate the property of interest, then it can be ruled out as a candidate explanation. If a given model has *no* parametric specifications that generate the phenomenon of interest, then the model as a whole is falsified. This in turn points to either (i) falsification of the theory that the model implements, or more likely (ii) a failure of one or more of the idealisations that are necessary in moving from a theory about the real world to an implemented

model. Conversely, if a model is not falsified, and has one or more parametric specifications that generate data that recapitulate some real-world phenomena, then we have a *possible* case of retrodiction, that is, a candidate explanation. Because of this complex interaction between necessary idealisations, auxiliary assumptions, in addition to the impracticality of exhaustively searching a parameter space, models rarely outright falsify theories, but rather serve the functions described in the following quote, from a recent book on modelling the origins of speech:

(1) They serve to evaluate the internal coherence of verbally expressed theories already proposed by clarifying all their hypotheses and verifying that they do indeed lead to the proposed conclusions (and quite often one discovers errors in the assumptions as well as in the conclusions, which need to be revised); (2) they serve to explore and generate new theories, which themselves often appear when one simply tries to build an artificial system reproducing the verbal behaviour of humans. (Oudeyer, 2006)

This thesis is in large part an exploration motivated by the first of these functions.

## 1.3 OUTLINE

**CHAPTER 2:** This chapter provides background material on several domains that this thesis cross-cuts. From a linguistic standpoint, I present those data that characterize the aspects of vowel harmony that typically interest phonologists, along with the theories and hypotheses (and some data) underpinning notions about coarticulation, phonologisation, and sound change. On the psychological side, I present exemplar models of categorisation, discussing their origins in psychological theories of categorization and contemporary applications to linguistic problems. Next, I discuss the computational and mathematical aspects of agent-based modelling and instance-based/lazy learning. The chapter finishes with discussion of some related work.

**CHAPTER 3:** I present LIBPHON, an instance-based model of phonological acquisition. After giving the details of the architecture and algorithmic aspects of LIBPHON, I show that it is capable of learning productive, generalised patterns of vowel harmony on the basis of limited exposure to acoustic formant trajectories. In addition to providing a novel account of vowel harmony, this chapter weighs against claims made in the generative literature that analogical mechanisms are too vague to be useful as theories of language acquisition and use (Chomsky, 1966), and that acquisition scenarios based on explicit storage of surface forms are incapable of gen-

eralising to novel inputs or creatively producing novel outputs (Chomsky, 1986).

**CHAPTER 4:** This chapter presents a model of sound change, in particular of the diachronic emergence of lexical vowel harmony. The set of simulations here give an existence proof for the viability of usage-based theories of sound change advanced by John Ohala and William Labov (*inter alia*). I will show, in particular, that patterns of vowel harmony can emerge diachronically from a stage lacking harmony via the mechanism of iterated learning in the context of biased transmission and acquisition.

**CHAPTER 5:** This chapter contextualises the work presented in chapter 3 and chapter 4. I discuss the bearing that this research has on phonology, both *as* and *in* cognitive science, and the compatibility of these views with currently and traditionally accepted notions of what phonology (*qua* field, rather than object of study) is about. In particular, I present a view of phonology divided along broadly Marrian lines (Marr, 1982), situating this and previous work within the hierarchy of levels of analysis.

**CHAPTER 6:** In the final chapter I review the key points of the thesis, discussing outstanding practical and theoretical questions, and suggest some potentially fruitful avenues for future explorations.



## CHAPTER 2

### BACKGROUND

#### 2.1 LINGUISTICS

##### 2.1.1 VOWEL HARMONY

Across a wide variety of languages and in virtually all language families, linguists have observed examples of vowel co-occurrence restrictions within specific phonological domains. These constraints on which vowels may appear together in a word are typically considered a unitary phenomenon, called *vowel harmony*. The vowels of a language with vowel harmony can be classified into disjoint sets, such that vowels from only one of the sets are found within some relevant domain, typically something like the phonological word, although this varies and harmony can occur

over larger and smaller domains. A standard example from the literature involves the front/back distinction in Finnish vowels (van der Hulst and van de Weijer, 1995):

	<i>output form</i>	<i>gloss</i>
a.	[tuhmasta]	‘naughty’ (elative)
b.	[tühmästä]	‘stupid’ (elative)

Table 2.1: Finnish backness harmony

The point to note in Table 2.1 is that elative case has two realisations—*-stä* and *-sta*—whose vowel backness depends on whether the stem has front {*ü, ä*} or back {*u, a*} vowels. Note that “vowel harmony” is used to describe two sets of phenomena that are perhaps best kept distinct: a generalisation over static (*viz.* non-alternating) lexical items—roots contain all front or all back vowels—and a generalisation that looks like a synchronic alternation, yielding patterns like those with the case suffix shown above.

Vowel harmony is found in many languages and across all major language families. The examples shown here barely scratch the surface of its complexity, and below I discuss exceptions to harmony processes, which can be either idiosyncratic (*e.g.* unregularised borrowings) or systematic.

### 2.1.2 NEUTRAL VOWELS

In most languages with vowel harmony, there are one or more vowels that systematically fail to harmonize. These are called *neutral vowels*, and

are typically further subclassified according to whether or not they induce alternations in nearby vowels.

The following data are from Tangale, a Chadic language with tongue root vowel harmony (van der Hulst and van de Weijer, 1995):

	<i>output form</i>	<i>gloss</i>
a.	[sebu]	‘look’ (imp.)
b.	[kɛnʊ]	‘enter’ (imp.)
c.	[peerna]	‘compelled’
d.	[pɛdna]	‘untied’
e.	[dɔbumgu]	‘called us’
f.	[dɪbnamɡʊ]	‘called you (pl.)’

Table 2.2: Tangale tongue root harmony. Item (e) is from Bakovic (2003).

In Table 2.2, items (a) and (b) show that the imperative morpheme surfaces with  $[\pm\text{ATR}]$  alternants, [-u] and [-ʊ] (see Lodge, 2009, on the problems with assuming a unified ATR feature), while (c) and (d) show that the past tense [-na] fails to alternate, hence is neutral. Item (f) is the crucial piece of data, showing that [-na] not only fails to alternate, but in fact induces  $[\text{ATR}]$  alternations on following vowels. This type of neutral vowel is called *opaque*.

In contrast to this, we also find neutral vowels which fail to harmonise, and do not induce alternations elsewhere, hence appearing “invisible” to the spread of harmony. In the Wolof system of tongue root harmony the vowels {i, u} fail to alternate (van der Hulst and van de Weijer, 1995):

	<i>output form</i>	<i>gloss</i>
a.	[toxileen]	“go & smoke !”
c.	[tɛkkilɛɛn]	“untie !”
b.	[seenuwoon]	“tried to spot”
d.	[tɛɛruwɔɔn]	“welcomed”

Table 2.3: Wolof tongue root harmony. Data from Archangeli and Pulleyblank (1994), with transcription normalised.

Note crucially that the suffixes that follow the non-alternating vowels agree in tongue root specification with the vowels preceding them. This type of neutral vowel is called *transparent*, and typically proves challenging for phonological theories. Mailhot and Reiss (2007) argue that these difficulties arise from a failure to recognise that the opaque/transparent dichotomy is more aptly cross-categorised as failure to harmonise, with or without concomitant induction of parasitic alternations elsewhere in the word.

The data described here are all demonstrations of so-called *stem-controlled* harmony in which root vowels are the determiners of vowel alternations in affixes. A second type of harmony, *dominant-recessive* (van der Hulst and van de Weijer, 1995, p.514), is characterized by feature-controlled agreement. In dominant-recessive systems, all vowels in a word alternate to agree with a particular feature, irrespective of whether it occurs in the stem or an affix. Standard examples of dominant-recessive language in the contemporary harmony literature are Turkana (Noske, 2000) and Nez

Percé (Aoki, 1966). The simulations described in this thesis only address stem-controlled harmony, but the account is in principle extensible to cover dominant-recessive cases as well.

### 2.1.3 CRITERIA FOR VOWEL HARMONY

Within the context of an account of the emergence of harmony from phonetic factors, how do we decide whether a language has vowel harmony? Three criteria have been acknowledged in the literature.

**LEXICAL STATISTICS** The proportion of harmonic roots in a language's lexicon may deviate significantly from the expected amount (as measured by some statistical or information-theoretic criterion) given its inventory of vowels (Goldsmith and Riggle, 2007; Denis, 2010), or may increase or decrease measurably on historical timescales.

**LOANWORD ADAPTATION** When disharmonic loanwords are borrowed into a language with vowel harmony, they may become harmonised over time or otherwise behave as expected given the language's harmony system (Zimmer, 1985; Kornai, 1990; Kertész, 2003). For example, Turkish has clearly identifiable sets of French and Arabic borrowings, from distinct historical periods. Both sets trigger suffix alternations, but only the historically older Arabic borrowings have become harmonised root-internally,

as well.

**SYNCHRONIC ALTERNATIONS** For many contemporary phonologists, productive and generalised surface alternations, such as the suffix alternations in the cases shown above, are the only true diagnostic of active harmony within a language:

“The only criteria that may play a role is the presence of two alternating sets of vowels in the inventory. When one set *induces the other to change*, vowel harmony exists in that language.”  
(Mahanta, 2007, p.14; emphasis added)

#### 2.1.4 COARTICULATION

To a first approximation, *coarticulation* is the label given to the relatively predictable effects that segments have on their neighbours in running speech. Coarticulation may affect adjacent consonants, as when an English speaker says [līmbejkən] for *lean bacon* (Kingston, 2007), or between vowels and consonants, as when an English speaker produces a relatively more palatalised [k] in *keep* versus *coop* or when a vowel is nasalised before a nasal consonant. Finally, it has been known since the work of Öhman (1966) that vowels may coarticulate with one another across intervening consonants, and in fact up to a distance of several syllables (Magen, 1997).

This vowel-to-vowel coarticulation is at the heart of one of the best-known explanations for the existence and typology of vowel harmony. Ohala (1993, 1994a) articulates the view (more recently taken up and expanded by Blevins 2004) that vowel harmony is a result of the phonologisation (see subsection 2.1.6) of this vowel-to-vowel coarticulation. In particular, he argues that harmony results when listeners are unable to “parse out” or compensate for the acoustic effects of distal segments (*viz.* neighbouring vowels) and misattribute contextual variation to the proximal segment.

Coarticulation has sometimes been explained in terms of articulatory undershoot and/or overlap and considered to be a largely physiological phenomenon resulting from the kinematics of speech (*viz.* of moving massive articulators to best hit a sequence of articulatory targets). However, there is a substantial body of empirical evidence that patterns of coarticulation and “perceptual compensation” for coarticulation (see subsection 2.1.5) are highly language-specific, in particular that anticipatory and perseverative coarticulation vary widely in degree across languages and that compensation for coarticulation is largely attuned to a particular language’s amount of coarticulation (Keating, 1985; Beddor et al., 2002, 2007). On the strongly internalist view of linguistic knowledge adopted in this thesis, this means that coarticulatory patterns are *speaker*-specific, hence, learned and essentially cognitive in nature. But if patterns of coarticulation are learned, there is no longer any reason to assume *a priori* that coarticulation is primarily a physiological effect, or perhaps even that word-internal coarticulation

exists at all:

If lexical representations are auditory, the details of word-internal coarticulation could be remembered as part of the lexical representation. (Coleman, 2000, p.110)

Where is coarticulation? Much of it is in the lexicon, the speaker's memory of what words sound like. (Coleman, 2000, p.117)

Clearly there are some aspects of coarticulation that are physiological, since speech *is* implemented in the movements of articulators with inertial mass. If the account being developed here, and entailed by the quotes above is right, a likely place for such effects would be between words (or at least words that do not frequently co-occur), or when a speaker is called on to produce a novel root-affix form:

Coarticulation at word-junctures might be more creatively planned during speech production [. . .] (Coleman, 2000, p.110)

The particular view of speakers' knowledge of phonology that I adopt here, namely that humans store "whole-word" exemplars in an unsegmented acoustic form, and more particularly that segments are not an ontologically distinct kind, forces me to the same conclusion.

Without these specific assumptions regarding the segmental nature of the phonological units, coarticulation is a non-



phenomenon. (Coleman, 2000, p.105)

The situation is complicated by the possibility of something like (diachronic) vowel merger. Investigations of V-to-V coarticulation seem to indicate that it is sensitive to the “crowdedness” of the vowel space, among other factors; languages with fewer vowels typically display greater amounts of coarticulation (Manuel, 1999). All else being equal, this predicts an increase in coarticulation post-merger, which on the Ohalian account would in turn lead to an increased likelihood of the development of vowel harmony. Conversely, many current phonological theories of vowel harmony make use of the theoretical device of “harmonic pairing” (Bakovic, 2003), which asserts that vowels in a language’s inventory that are not “paired” for a harmonic feature tend to be neutral.<sup>1</sup> These accounts hence predict that the result of merged vowels (assuming the merger is along the relevant dimension) should be neutral, and hence not affect a language’s overall patterns of harmony, either synchronically or diachronically.

### 2.1.5 PERCEPTUAL COMPENSATION

Humans can parse out, or perceptually compensate for, coarticulation. Beddor et al. (2002) elicit success rates for English and Shona in a some-

---

<sup>1</sup>Mailhot and Reiss (2007) argue against the usefulness of the notion of harmonic pairing, given that researchers are typically inexplicit about whether the pairing is meant to apply at the surface or underlying levels. Moreover, the notion of a vowel inventory loses coherence with the introduction of underspecification.

what complicated experimental setting, playing bisyllabic nonce words with syllables cross-spliced into coarticulatorily-inappropriate contexts. Relatedly, and relevantly for the work presented here, Ohala (1994a) suggests that high (*esp.* high front) vowels are perceptually salient and hence particularly easy for the listener to parse out, which could account for the fact that transparent vowels in harmony systems tend to be high front vowels.

### 2.1.6 PHONOLOGISATION

Historically, *phonologisation* was a term used to describe the diachronic process whereby linguistic variation under physical/physiological (*i.e.* “phonetic”) control came to be under cognitive (*i.e.* “phonological”) control. The term was introduced by Jakobson<sup>2</sup> and was reintroduced to the generative literature in Hyman (1972). The following quote is illustrative of this view:

A universal phonetic tendency is said to become “phonologised” when language specific reference must be made to it, as in a phonological rule. (Hyman, 1972, p.170)

An example is vowel nasalisation in standard varieties of American English (*e.g.* *pit*~[pɪt] *vs.* *pin*~[pɪ̃n]). These vowels are typically taken to be

---

<sup>2</sup>But it was discussed as a mechanism of sound change as early as Baudouin de Courtenay (1895).

“more nasal” than would be expected given the amounts of anticipatory coarticulation that are typically found with a following nasal, and so a rule is posited that adds a nasal feature to the vowel. In some cases, the trigger for the alternation is lost for independent historical reasons while the target segment retains its changed form. Hyman calls these cases *phonemicisation*, and cites French nasal vowels as an example.

My focus here, with respect to acquisition, is on a level of analysis at which traditional linguistic constructs are unavailable, and so this definition is in need of revision. A relatively theory-neutral description of phonologisation is *the cognitive encoding by acquirers of systematic variation in the primary linguistic data*. Given the particular instance-based view to be adopted in this thesis, this view entails that “phonologisation” has no standing as a theoretical construct, or rather that it is non-distinct from the general mechanism of phonological acquisition, ongoing in all speakers at all times, without endpoint.

### 2.1.7 THE ORIGINS OF PHONOLOGICAL ASSIMILATION

Ohala (1993) provided the standardly accepted answer to the question of how assimilatory phonological phenomena originate: gradient patterns of coarticulation are misperceived and/or misparsed by acquirers, and over time (possibly multiple generations) become phonologised as categorical patterns of phonological assimilation. For the particular case of vowel

harmony, Ohala (1994a) suggested that the exact same scenario plays out in the domain of vowel-to-vowel coarticulation. A straightforward corollary of this is that “vowel harmony” is simply a variety of assimilation, and consequently does not merit any privileged ontological or theoretical status in synchronic theories of phonological knowledge. Maillhot and Reiss (2007) come to the same conclusion working within a more traditional generative-formalist framework.

One of the chief aims of the work presented here is to examine what means there are, if any, of verifying or supporting this type of diachronic explanation of a synchronic pattern. In order to have a viable Ohalian explanation of this type, at minimum the following need to be given:

1. a demonstration of synchronic variation in production (Ohala, 1989),
2. a demonstration that this variation is detectable by listeners (Ohala, 1981),
3. a relatively worked out model of synchronic linguistic knowledge,
4. a relatively worked out model of language acquisition, and finally
5. *a demonstration that the previous items can bring about the phenomenon under consideration, given sufficient time*

If Ohala’s (1981; 1989) claims are correct about the role of the listener in sound change, and about sound change being a product of synchronic variation, then item 1 and item 2 acknowledge that phonologisation is

essentially a species of Neogrammarian sound changei (Hale, 2007), and items 3 and 4 are simply requirements on the explicitness of auxiliary assumptions. These are relatively uncontroversial, and are the bread and butter of experimental and theoretical phonologists. On the other hand, item 5 leads to difficulties. There is no obvious way to verify or test the diachronic dimension which is crucial to this kind of explanation. To be sure, one can make and record some predictions and trust that their (dis)confirmation will be followed up on by future generations, but this is a rather unrewarding way of doing research. Moreover, it is exceedingly unlikely that any predictions specific enough to be testable would ever survive falsification (see previous discussion on this point), given the amount of uncontrollable factors, *e.g.* patterns of connectivity and communication in social networks, language contact situations, *etc.* (but *cf.* Niyogi (2006) and references discussed in Hruschka et al. (2009) for recent attempts to address some of these issues). Of course, rather than making predictions about specific occurrences of change, diachronic explanations are more standardly taken to make *typological* predictions and retrodictions, which are in principle open to immediate verification. In other words, if a particular change is predicted to be likely or frequent, one assumes that its outcome will be typologically well-represented. Of course, typological data are equally likely to be subject to noise and extraneous factors, perhaps even more so, given the difficulty of controlling for factors like availability of resources for studying certain language groups,

language extinction rates, *etc.*

### 2.1.8 ON LANGUAGE AND LANGUAGE CHANGE

#### *I*-LANGUAGE VS. *E*-LANGUAGE

Chomsky (1986) introduced the notion of *I*-language (*cf.* Jackendoff, 2002, for cogent discussion). It is the set of cognitive entities and operations that speaker-hearers (*mutatis mutandis* signer-seers) have that enables flexible, productive, generalised use of language in production and comprehension. The *I*- denotes “internal, individual, intensional,” the first two emphasising that *I*-language is a property of individuals, instantiated within their minds/brains, and the last emphasising the “summary” nature of linguistic knowledge, *viz.* that it is a *grammar*, a system of rules and/or constraints sufficient to generate/recognise the infinite set of sentences that adult native speakers are claimed to be in command of. Crucially, *I*-language is distinct from Chomsky’s notion of *competence*, which he introduces as a theoretical entity, described as the *I*-language of an idealised speaker-hearer in a linguistically homogeneous environment, subject to no externally-caused deviances (*i.e.* embedded in a perfectly monolingual environment, having no memory problems, not subject to slips of the tongue, *etc.*).

The point of introducing *I*-language is to differentiate it from what Chomsky calls *E*-language, which is “external” and “extensional.” *E*-language is

typically defined as one of: (i) everything linguistic that is not *I-language*, or (ii) the shared/common language of a linguistic community (essentially de Saussure's *langue*). One of Chomsky's 1986 core claims is that *E-language* on either of these construals is not coherent enough to be the object of scientific investigation. The definition of competence, however, strikes me as precisely an attempt to delineate that subset of *E-language* which might be amenable to formalisation.

The instance-based approach to language that I adopt here is intermediate between these positions. It is *internal* and *individual*, preserving what I think are the crucial insights in Chomsky's original discussion, namely that humans have a cognitively encoded body of knowledge that permits the flexible, productive, creative use of language. Different, though is an *extensional* stance, in the sense that language is represented as a set—not necessarily structured—of memorised previously encountered forms. Of course, in addition to these memorised linguistic experiences, there remains a need for some analogical mechanism enabling generalisation from remembered forms, both in production and comprehension. I discuss this mechanism and illustrate its implementation in detail below.

### ON LANGUAGE CHANGE

Lass (1997), in a discussion of the ontological status of “change” for the historical linguist, distinguishes *translative* change, in which a single en-

tity or natural kind becomes something different from what it was at a chronologically prior point, from *substitutive* change, in which there is a sequence of entities, states or natural kinds chronologically juxtaposed, none of which individually changes. Both of these positions have been adopted at various times as characteristic of language change, *cf.* Lass (1997) for the former and Janda and Joseph (2003) and Hale (2007) for the latter. While this distinction probably makes no practical difference (*viz.* with respect to the end product of change), as Lass points out:

[O]ntological commitments [...] can generate heuristic metaphor and 'vision'; these may play a vital role in giving shape to one's imaginings, and from there can lead to particular more sharply focused theories or models. (Lass, 1997, p.280)

Consider, for example, whether a notion like "push/pull chain" in accounts of vowel shifts would have arisen on a substitutive view of change.

One of the foci of this thesis is on individual-level events as drivers of (one kind of) phonological change. In a sense, then, both the translative and substitutive views of change are pertinent to the account to be given here; on the first hand, I am arguing that change is initiated in the biased transmission and acquisition mechanisms that individuals are disposed of, and by keeping my focus suitably narrow I conceptualise change as a sequence of individual grammars (*I*-languages) juxtaposed across generations. On the other hand, I am concerned with the distribution of particular forms



(those exhibiting some degree of vowel harmony) across a population, over time. Especially given the understanding that “generations” are a convenient fiction, there does not seem to me to be a clear sense in which this latter concern can be usefully viewed in substitutive terms, distributions being atemporal descriptive entities.

### 2.1.9 THE ACTUATION PROBLEM

In a landmark paper laying out a series of foundational questions for historical linguistics, Weinreich et al. (1968) posed the so-called *actuation problem* as a core question in the study of language change. The question essentially asks why a particular instance of language change happens when and where it does:

What factors can account for the actuation of changes? Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times? (Weinreich et al., 1968, p.102)

For even when the course of a language change has been fully described and its ability explained, the question always remains as to why the change was not actuated sooner, or why it was not simultaneously actuated wherever identical functional proper-

ties prevailed. The unsolved actuation riddle [...] creates the opposite problem—of explaining why language fails to change. (Weinreich et al., 1968, p.112)

The actuation problem is posed as a riddle in particular for theories of sound change that are grounded in individual-level “least-effort” drives (e.g. the theories of Hermann Paul), but *ceteris paribus* they would seem to apply to any theory of sound change based in the phonologisation of phonetic tendencies. In chapter 4 I return to the actuation problem and show that it poses less of a problem to the account developed here than might be expected.

## 2.2 INSTANCE-BASED APPROACHES TO COGNITION AND PERCEPTION

Instance-based models of cognition, also called exemplar-based, memory-based and case-based models, have their (modern) origins in psychological theories and models of perceptual categorisation and episodic memory (cf. Medin and Schaffer, 1978; Hintzman, 1986; Kruschke, 1992; Tulving, 1972). The earliest explicit discussion is in Semon (1921), a theory of memory that anticipates many of the features of contemporary exemplar models. The core properties characterising instance-based models of cognition are:

1. explicit storage of input/training data, with associated categorical information, and
2. processing of novel inputs based on similarity to stored instances, where similarity is a function from the parametric space that the instances inhabit to some totally ordered set.

In the following sections I present a few of the better-known instance-based models in psychology and linguistics, and their analogues in machine learning.

### 2.2.1 EXEMPLAR MODELS IN PSYCHOLOGY

#### MEDIN & SCHAFER

Some years after the introduction of prototype models into the categorisation literature in cognitive psychology (*cf.* section 5.1 for discussion), Medin and Schaffer (1978) noted that many of the empirical effects previously attributed to the similarity of stimuli to internal prototypes (*e.g.* typicality and reaction time gradients) were equally well explained by assuming that people mentally store all tokens of perceptual experiences and classify stimuli on the basis of similarity to the entire group of remembered episodes. They introduced the *Context Model* of classification, the first exemplar-based psychological model with an explicit algorithm for computing similarity, and successfully accounted for a variety of phe-

nomena previously thought to be explainable only on the basis of mental prototypes.

### NOSOFSKY'S GENERALIZED CONTEXT MODEL

Nosofsky (1986), building on the work of Medin and Schaffer (1978), introduced the *Generalized Context Model* (GCM), which according to many authors is the most successful exemplar-based model of categorisation to date (cf. Murphy, 2002, *inter alia*). The GCM computes the posterior probability that a person will assign a particular category label to a given stimulus by calculating the feature-wise similarity of the stimulus to all exemplars within a category, divided by the feature-wise similarity to all exemplars across all categories, as shown here (omitting some additional parameters not material to the discussion):

$$P(C_m|x_i) = \frac{\sum_{x_j \in C_m} \pi(x_i, x_j)}{\sum_{x_k} \pi(x_i, x_k)}$$

where  $C_m$  is a particular category label,  $x_i$  is the input data token, and  $\pi(x_i, x_j)$  and  $\pi(x_i, x_k)$  are functions of the distance between the input datum and other tokens in the parametric instance-space.

**HINTZMAN’S MINERVA2**

Hintzman’s (1986) MINERVA2, another well-known exemplar model, operates by encoding each perceptual episode as a fixed-width ternary vector in long-term memory, with values in  $\{-1, 0, +1\}$ . When a new stimulus—a “probe” in Hintzman’s terminology—is presented to MINERVA2, its features are compared to all memory traces and a multiplicative similarity measure—essentially a normalised dot product—is derived:

$$S(p, x_i) = \frac{1}{N} \sum_{j=1}^n p_j \cdot x_{i,j}$$

where  $p$  is the probe,  $x_i$  is the  $i^{\text{th}}$  memory trace, and  $N$  is the number of nonzero elements in the element-wise product of  $p$  and  $x_i$ . Each trace that has features in common with the probe becomes “activated” to a degree proportional to its similarity to the probe. All traces with activated features are summed into a transient summary representation—an “echo”—which is returned *e.g.* as a paired associate or category judgment.

**2.2.2 EXEMPLAR MODELS IN LINGUISTICS**

In linguistics, exemplar models have only become popular in the last 15 years, first being proposed to account for data from speech perception experiments which showed that some aspects of linguistic knowledge are

rooted in experiential episodes (*cf.* Goldinger, 1996), and subsequently being adopted into phonetics (*e.g.* Johnson, 1997b; Hawkins, 2003; Coleman, 2002) and phonology (*e.g.* Pierrehumbert, 2001, 2003; Wedel, 2004a).

The exemplar models implemented to date in phonetics and phonology have largely focused on perception (*e.g.* speaker normalisation in Johnson 1997b), or on segment-internal diachronic processes (Pierrehumbert, 2001; Ettlinger, 2007; Wedel, 2007), leaving the types of phenomena that typically interest “traditional” phonologists (*viz.* productive, general “processes”) comparatively neglected, although Kirchner and Moore (2009) demonstrate what is arguably a synchronic model of spirantisation.

### ANALOGICAL MODELING

Skousen (1989) is the earliest clear example of a modern instance-based approach to language modelling. Although Skousen and his students and colleagues have pursued this approach for over 20 years, analogical modelling (AM) has had a limited impact on the broader linguistic and cognitive scientific community. In part this may be because of the idiosyncratic terminology that Skousen employs, along with the computational intractability of his earlier models. AM differs from instance-based models as defined thus far in several ways. On the first hand, it has nearly always been applied to phenomena defined over discrete, in fact segment-level, units. Additionally, the selection of the analogical set—the collec-

tion of tokens over which analogies are computed in categorising novel inputs—is not based on a distance-based calculation, but rather involves a lattice-theoretic comparison of subsets of features that is exponential in computational complexity. AM has had some success modelling particular linguistic patterns (Skousen et al., 2002), although it has not been extensively applied in the domain of phonology.

## **XMOD**

Johnson (1997a,b) essentially introduced instance-based models to the community of phoneticians and phonologists. The initial application of his model, XMOD—a modified version of Nosofsky’s GCM—was to speaker normalisation in speech perception, and in particular Johnson argued that on an exemplar view, no active mechanism of normalisation is needed, provided listeners store speech tokens with indexical information, in addition to parametric phonetic information. Johnson also applied XMOD to a demonstration of the emergence of segment-like and syllable-like units from auditory spectrograms of continuous speech. By tracking activation (*viz.* similarity) dynamics across tokens of different words unfolding in time, Johnson showed that patterns of coherence could be seen on time-scales similar to those of segments and syllables. More recently, he has given some historical context to exemplar models in phonology, and argued for particular choices of input dimensionality and parametrisation

(Johnson, 2007). I return to this in section 3.1.

### **PIERREHUMBERT 2001 *et seq.***

Pierrehumbert (2001) was the first to take up the instance-based approach to phonology after Johnson's papers, and the first to give an explicit *production* algorithm, prior linguistic instance-based models having been strictly classification-based. To generate an output, a token exemplar from a category is randomly selected, and an analogical set is constructed based on a local neighbourhood. Pierrehumbert's model, in contrast to those discussed above, explicitly restricts itself to a subset of the stored exemplars in computing classifications and outputs by considering a fixed-radius neighbourhood around the location in parametric space of the exemplar to be classified or produced. Classification essentially uses the *k*-nearest-neighbours algorithm. Pierrehumbert uses her model in a production-perception feedback loop to model a hypothesised diachronic leniting sound change.

### **PEBLS**

Kirchner and Moore (2009) present PEBLS, an exemplar-based production model that uses real speech tokens. Rather than using auditory spectrograms, like Johnson, they transform raw acoustic data into *mel frequency*



*cepstral coefficients* (MFCC), a psychoacoustic representation that is common in the speech processing literature (Deng and O'Shaughnessy, 2003). PEBLS uses a variant of *dynamic time warping* to define similarity across tokens and is able to model synchronic intervocalic spirantisation; if populated with a set of spoken tokens of [axa] and asked to produce [aka], the velar consonant will show some degree of frication in the output. As of the time of writing, PEBLS is not able to generalise outside of a category of exemplars (*e.g.* in the previous example, if asked to produce [ake], PEBLS will not fricate the velar).

### THEORETICAL MODELS

In addition to the implemented instance-based models of phonology discussed above, several linguists have argued for and proposed similar theoretical models. Joan Bybee has argued strongly for a lexicon that is rich in phonetic detail, including frequency information, on the basis of the role that word-specific frequencies play in diachronic processes (Hooper, 1976; Bybee, 2001). Similarly, Robert Port has argued against the discrete, low-dimensional representations employed by traditional schools of phonological theory, arguing instead for richly-detailed lexical representations embedded in high-dimensional acoustic spaces (Port and Leary, 2005; Port, 2007b).

Finally, Cole (2009) presents an exemplar-based account of vowel harmony.

Cole motivates her model from empirical inadequacies of traditional generative accounts of vowel harmony—mostly pertaining to overgeneration of predicted patterns of harmony—and artificial grammar learning experiments which seem to demonstrate functional benefits of vowel harmony. The model she describes assumes the existence of both phone-sized and word-level units, and the productivity and generality of vowel harmony emerges from *resonant* interactions between the two (Grossberg, 2003). The account is implicitly perception/classification-oriented and it is unclear whether or not it could be straightforwardly extended to production of novel harmonic forms.

#### A POTENTIAL CRITICISM

Albright and Hayes (2003) compare the performance of their *Minimal Generalisation* rule-based learner and an exemplar-based “analogical learner” in acquiring a means of generating English past tense forms, both regular and irregular. Their analogical learner is a version of the GCM (*cf.* 2.2.1). The input data in this domain are massively variable, with many exceptional forms and subregularities, and constitute a particularly challenging test for any computational learner. Albright and Hayes show that their version of the GCM generally underperforms in comparison to their rule-based learner. They point out a particular type of error that the analogical learner makes in choosing which past tense allomorph to use in a “wug

test” type scenario, stating:

Locating the final consonant to determine the correct ending is a canonical case where structured similarity is required: the past tense allomorph depends solely on the final segment of the stem, in particular on just a few of its features. Our analogical model, however, is inherently unable to focus on these crucial structural elements. Instead, it gets distracted by variegated similarity, and makes wrong guesses. For instance, for the existing verb *render*, the analogical model guesses *\*renderèd* [rɛndərəd], based largely on the following analogical set (the ten most similar forms): *rend*, *end*, *rent*, *vend*, *raid*, *fend*, *mend*, *tend*, *round*, and *dread*. These stems bear an irrelevant similarity to *render*, which (in this case) suffices to outweigh the influence of legitimate model forms like *surrender*. (Albright and Hayes, 2003, p.151)

On this basis, Bruce Hayes (personal communication) has expressed doubts about the ability of any “purely analogical model”, in particular LIBPHON, to successfully capture the empirical facts of interest.

Claims about the deleterious effects of variegated similarity on producing correct outputs are particularly important to address in the context of a model like LIBPHON, which is to some extent expressly about the non-necessity of traditional structure-bound notions of similarity. Unfor-

Unfortunately, there are several factors contributing to the incommensurability of Albright and Hayes’s model with LIBPHON: they use symbolic/segment-based representations, calculating similarity according to string-edit distance and a natural-class based feature-counting metric, their lexicon is type- rather than token-based, and there is no “semantic” category information available to restrict the domain of the analogical computation. Hence, in selecting a form to output, their model compares categories that are never in competition in LIBPHON. As a final note, from a purely empirical point of view, it strikes me that there is a problem with their similarity computations (in particular the alignment algorithm) if, for example, it finds *render* to be closer to *vend* than *surrender*. In any case, the data used in these experiments have been made available, and an investigation of LIBPHON’s performance on this task seems a worthwhile avenue for further work.

## 2.3 COMPUTATIONAL AND MATHEMATICAL BACKGROUND

### 2.3.1 TYPES OF MODELS

Modelling strategies can be loosely classified as either *analytic* or *synthetic*. Analytic models are based on closed-form mathematical equations, such

as systems of differential equations, which are solved numerically by computer. In a language modelling context, analytic models are often focused on population-level properties, *e.g.* the proportion of a speech community adopting a particular variant of some linguistic form. A book-length treatment of this kind of work is (Niyogi, 2006). Synthetic models, on the other hand, are primarily concerned with modelling individuals; population-level properties, to the extent that they are of interest, are viewed as being emergent from the inter-individual dynamics. These kinds of models are more commonly called *agent-based* or *multi-agent* models. The work described in this thesis is an example of synthetic language modelling.

Epstein (2006), in a discussion of agent-based modelling in the social sciences, cites the following properties as characteristic of (though not constitutive of) synthetic models:

- Heterogeneity: individuals are explicitly represented, and may vary in myriad ways
- Autonomy: there is no global (“top-down”) control on agents’ behaviour (although there may be feedback between macro- and micro-level properties)
- Explicit space: agents are embedded in a “space”, which can be more or less abstract (*e.g.* artificial world with implemented physics *vs.* closed 2-d torus of cells)

- Local interactions: agents typically interact with nearby neighbours in their space
- Bounded rationality: agents do not have perfect information about their world and have finite/limited computing resources

### 2.3.2 INFORMATION FLOW

Synthetic models can be further subclassified according to constraints on the flow of information between agents. In a *horizontal* model, any pair of agents can interact with one another, and all agents can update their internal state. In these types of models, there are no explicit targets towards which agents' internal models converge. Instead, they generally display *self-organisation*, with order spontaneously emerging from a randomly-specified initial global state. In a *vertical* model, there are restrictions on which pairs of agents can interact and which agents have modifiable internal states.

Kirby (1996) introduced and popularized linguistic agent-based models with vertical information flow as *iterated learning* models. In an iterated learning model, the population of agents is partitioned into two disjoint classes, one with fixed internal state (modelling “adults” or “teachers”) and the other with modifiable internal state (modelling “children” or “learners”). Agents may only communicate across classes, and most typically,

children are listeners and adults speakers. The adult grammars serve as approximate targets toward which the child grammars are meant to converge. Upon convergence, or after some predetermined amount of time, the adults are replaced by the children, whose internal states become fixed, a new generation of children is introduced, and the process is repeated. There are two potential drivers of change in these models: noisy data transmission and the information bottleneck that obtains when learners are exposed to only a subset of the data.

### 2.3.3 INSTANCE-BASED MODELS AND LAZY LEARNING

Instance-based models, introduced to the machine and statistical learning communities by Fix and Hodges (1951), are well-known in the machine learning literature and many of their formal and theoretical properties—in particular their error bounds—have been derived (Cover and Hart, 1967). Recall the basic properties of exemplar models, repeated here from section 2.2:

1. explicit storage of input/training data, with associated categorical information if it is available, and
2. categorisation of novel inputs based on similarity to stored instances, where similarity is typically a function of distance in the space that the instances inhabit.

A third property, implicit in many of the psychological models considered above but rarely commented on, is that instance-based models are characterised by *deferred evaluation*; all computation is delayed until the system is called on to classify or produce a datum. Contrast this with a typical connectionist system/artificial neural network (Hertz et al., 1991): a collection of simple computing units equipped with activation functions that are linked by weighted connections. On the basis of training data, the system modifies its weights using some (typically error-driven) algorithm. Once a convergence criterion has been reached, the training data are discarded, and the neural network's weights embody an intensional representation of a function defined over the entire input parametric space.

Taken together, these properties point to a more perspicuous naming scheme for the method under consideration. In some sense, *all* learning algorithms are “instance-based”, in that they use input data to either learn an approximation to a function (supervised learning), or to discover latent structure in the input (unsupervised learning). The property of deferring generalisation until query time emerges as the key aspect differentiating instance-based models from the kinds of learning algorithms more generally familiar in the cognitive sciences. In the machine learning literature, this is known as *lazy learning*, to be contrasted with *eager* methods (Aha, 1997), in which some generalised representation is updated as each input datum is received, *e.g.* the connectionist system described above, or standard decision tree or rule-induction algorithms. Lazy learn-



ers typically create an approximation to a target function that is valid only in a *local* neighbourhood of some query point, rather than constructing a global approximation designed to be applied over the entire instance space. For a given parametrisation, lazy learners are therefore able to encode a more complex hypothesis space than eager learners (Mitchell, 1997), as illustrated by so-called “gang effects”. In order to avoid overfitting the training data (and hence sacrificing generalisation), eager methods use various means to introduce bias, and hence smooth over outliers. With a lazy learner, however, a cluster of similarly-categorised outliers can sway a classification or generation algorithm to fit local irregularities, without sacrificing any power to generalise elsewhere in the space of inputs.

#### **LAZY LEARNERS AND DENSITY ESTIMATION**

Ashby and Alfonso-Reese (1995) were the first to explicitly draw a parallel between categorisation and probability density estimation. In particular, they noted that standard exemplar models from the psychological literature (*e.g.* Nosofsky’s GCM) are essentially variations on a family of density estimation methods well-known in the statistical learning literature; non-parametric kernel-based density estimators (Schölkopf and Smola, 2002). Kernel methods work by defining distance-based functions, *kernel functions*, that define similarity gradients in the local neighbourhoods of stored data points (an intuitive conceptualisation has each data point replaced by

a Gaussian “bump”). Different kernel functions define different neighbourhoods (*e.g.* selecting a fixed volume around a point in the instance space or letting the volume grow to include a fixed number of neighbours) each of which in turns leads to different shapes of similarity gradients, generalisation properties and convergence guarantees. Many of the best performing machine learning algorithms today are kernel-based. Jäkel et al. (2008) give an excellent overview of the relation between kernel methods and exemplar models, relating the properties of specific models to different types of kernel-based estimators and discussing generalisation properties.

## 2.4 PREVIOUS/RELATED WORK

### 2.4.1 FORMAL ACCOUNTS OF VOWEL HARMONY

I discuss here a handful of the previously-proposed theoretical accounts of vowel harmony. The literature is vast, so I have restricted myself to a sampling that reflects some of the main approaches from the history of generative phonology.

#### LINEAR ANALYSES

**ROOT-MARKER THEORY** Lightner (1965) gave the first generative account of vowel harmony. On this account, roots in the lexicon are divided into

two categories on the basis of an arbitrary diacritic,  $[\pm Z]$ . A UG convention distributes this feature to all segments, then a phonological rule translates the  $[\pm Z]$  specification into a phonological one, *e.g.*  $[\pm \text{ATR}]$ . On this view, harmony actually arises from the convention distributing lexical features across segments, rather than from a phonological rule. This approach evidently runs aground on disharmonic stems, especially in forms like Akan [ɔ-kari-i], where the  $[-\text{ATR}]$  prefix harmonizes with the leftmost vowel in the root, while the  $[\text{ATR}]$  suffix harmonizes with the rightmost vowel. There is no single diacritic feature that can be assigned to the root that can subsequently result in prefixes and suffixes with opposite specifications for the harmonic feature.

**MORPHEME-STRUCTURE CONSTRAINTS** Kiparsky (1968) proposed to account for vowel harmony with a different two-part process. *Morpheme structure* conditions—essentially descriptive generalisations over the lexicon—handled root-level harmony, and an assimilatory phonological rule assigns the appropriate value to alternating affixes. Schematically, this looks like:

$$1. C_0 \begin{bmatrix} V \\ \alpha \text{ATR} \end{bmatrix} \cdots C_0 \begin{bmatrix} V \\ \alpha \text{ATR} \end{bmatrix} C_0 \quad (\text{MSC})$$

$$2. V \longrightarrow \alpha\text{ATR} / \left\{ \begin{array}{l} \text{--- } C_0 \left[ \begin{array}{c} V \\ \alpha\text{ATR} \end{array} \right] \\ \left[ \begin{array}{c} V \\ \alpha\text{ATR} \end{array} \right] C_0 \text{---} \end{array} \right. \quad (\text{assimilatory rule})$$

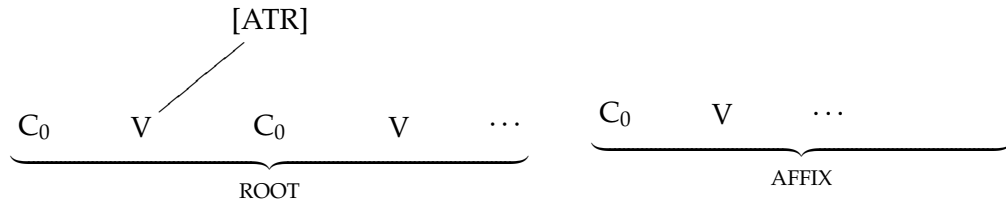
where  $C_0$  stands for a string of zero or more consonants, and the  $\alpha$ -notation denotes agreement irrespective of feature specification. The MSC-based account runs into problems with ordering paradoxes (*e.g.* in Turkish where a palatal umlauting process interacts with labial harmony). In addition, some researchers were uncomfortable with the separation of root-level and affixal harmony into separate processes, given the absence of languages with harmonic roots and non-harmonising affixes.

### NONLINEAR ANALYSES

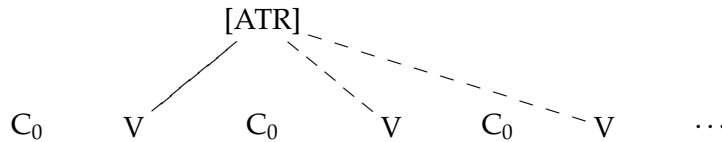
Autosegmental approaches to vowel harmony (*cf.* Clements, 1980) retained the intuition of the root-marker theory that harmony is somehow a property of entire root morphemes, rather than any particular root vowel. By eliminating the requirement of one-to-one correspondence between features and segments that characterised previous linear accounts, an autosegmental analysis allows root-level harmony to be directly specified in terms of phonological features, rather than arbitrary root markers. Segments that alternate are underlyingly unspecified for the harmonic feature, which they receive by autosegmental spreading (subject to a ban on cross-

ing association lines, which deals with opaque vowels). Schematically:

1. Root with ATR autosegment associated to leftmost root vowel



2. Phonological feature spread via autosegmental association



### SEARCH AND COPY

In Mailhot and Reiss (2007), Charles Reiss and I essentially formalised the account of Kiparsky (1968) in terms of a procedural **SEARCH** operation that is parametrised for direction, and feature-specification of target and trigger, and a simple **COPY** operation. Following Raimy (2000), we took phonological representations to be flat strings of phonological feature matrices with an explicit ordering relation. Our account of Turkish assumed the following schematic underlying structure for Turkish genitive plurals, which show both kinds of harmony found in Turkish (the plural morph is  $lVr$  and the genitive is  $Vn$ , where  $V$  is an alternating vowel):

$$\text{ROOT} \rightarrow \text{l} \rightarrow \begin{bmatrix} \text{V} \\ -high \\ -round \end{bmatrix} \rightarrow \text{r} \rightarrow \begin{bmatrix} \text{V} \\ +high \end{bmatrix} \rightarrow \text{n}$$

and proposed the follow rules:

a. Palatal harmony

- i. From  $[\zeta : \text{v}]$  SEARCH left for  $[\gamma : \alpha_{\text{BACK}}]$
- ii. COPY  $[\alpha_{\text{BACK}}]$  to  $\zeta$

b. Round harmony

- i. From  $[\zeta : \text{HI}]$  SEARCH left for  $[\gamma : \alpha_{\text{ROUND}}]$
- ii. COPY  $[\alpha_{\text{ROUND}}]$  to  $\zeta$

SEARCH is initiated from underlyingly unspecified vowels and proceeds until a target satisfying the stopping criterion is found, whereupon the COPY operation supplies the underspecified trigger with the missing feature (we were inexplicit about whether this is actual copying, or something more like the structure sharing in *e.g.* HPSG representations). This purely phonological account eliminates the need for morpheme-structure constraints, and in addition handles the umlauting interaction with labial harmony straightforwardly.

**OPTIMALITY THEORETIC ANALYSES**

There are, broadly speaking, two competing views of vowel harmony in the Optimality Theory (Prince and Smolensky, 2004) literature.

The chronologically older approach treats harmony as deriving from alignment constraints on vocalic features (Smolensky, 1993; Cole and Kisseberth, 1994; Pulleyblank, 1996, *inter alia*). On this approach, which at least implicitly preserves much of the representational machinery of autosegmental accounts, a violation is assessed whenever a candidate has a conflicting feature specification between a vowel at one edge of the designated domain and the other edge of the designated domain.

The newer, agreement-based, approach (Bakovic, 2000; Krämer, 2003) views harmony as an effect of the interaction between (input-output) faithfulness constraints and (syntagmatic) correspondence constraints. In particular, an AGREE constraint (not always called by that name) enforces featural agreement in adjacent vowels, while standard FAITH constraints preserve underlying features. Relativised to particular features and to stem *vs.* affix, the interaction of these constraints derives many of the phenomena traditionally referred to as vowel harmony, although transparent neutrality requires a certain amount of *ad hoc* machinery (*e.g.* specially targeted constraints).

### 2.4.2 COMPUTATIONAL ACCOUNTS OF VOWEL HARMONY

#### ELLISON 1992

Ellison (1992) shows that the distributional restrictions on harmonic classes of vowels can be modelled by 2-state nondeterministic finite automata (N2A). This allows him to recast the harmony acquisition problem as finding the N2A which best encodes a given harmonic corpus, where “best” is operationalised in terms of a *minimum description length* (Grünwald, 2007, though Ellison does not explicitly use that name) driven search over the subset of the space of N2As that can be interpreted as allowing basic harmony or neutrality.

#### GOLDSMITH ET AL. 2007/2009

John Goldsmith and colleagues (Goldsmith and Riggle, 2007; Goldsmith and Xanthos, 2009) have pursued probabilistic and information-theoretic approaches to the detection and learning of vowel harmony. Goldsmith and Xanthos (2009) train a 2-state Hidden Markov Model (Manning and Schütze, 1999) on Finnish vowel sequences and show that the HMM that assigns greatest likelihood to the corpus learns to separate the vowels into front, back, and neutral categories (more specifically, front and back vowels are segregated by HMM state and neutral vowels have roughly equal probabilities for both states). Goldsmith and Riggle (2007) use a bigram-



based approach (Manning and Schütze, 1999) optimising pointwise mutual information to model vowel harmony in Finnish, and show that a naive application of this approach to a “vowel-tier” representation of Finnish does not capture harmony significantly better than a bigram approach over the unprocessed corpus (*i.e.* including vowels and consonants). The best performance is obtained with a two-tier model combining the plain bigram model and a distribution over non-adjacent vowels.

### KARTTUNEN 1993

Karttunen (1993) discusses an approach to phonology initiated by Johnson (1972) and Kaplan and Kay (1981), which is grounded in the realisation that the relation between lexical and surface forms is *regular* (in the formal sense). Thus, lexicon-to-surface mappings are expressible as finite-state transducers or in a declarative rule formalism that resembles traditional rewrite rules, but with 4 different relational connectors (expressing different relational modalities) and in which contextual constraints can apply either at the lexical or the surface level.

In the context of this discussion Karttunen gives a finite-state analysis of Turkish vowel harmony. The (mostly intuitive) abbreviations for particular vowel sets given here make the formulation of the relevant rules less cumbersome:

RndVowel	=	o u ö ü	HiBackVowel	=	I u
BackVowel	=	a I o u	LowBackVowel	=	a
HighVowel	=	i I ü u	HiHrmVowel	=	I ε
HiRndVowel	=	ü u	LowHrmVowel	=	E

where  $\varepsilon$  denotes the empty (unrealised) symbol. In the Turkish system (to a first approximation), vowels harmonise with respect to backness, and alternating high vowels also agree with respect to roundness. The relevant harmony rules are then:

### Rounding

$$\text{HiHrmVowel}:\text{HiRndVowel} \Leftrightarrow :\text{RndVowel} :C^* \text{ \_\_\_}$$

### Backness

$$Vx:Vy \Leftrightarrow :\text{Backvowel} :C^* \text{ \_\_\_};$$

where  $Vx$  in (HiHrmVowel LowHrmVowel) matched with

$Vy$  in (HiBackVowel LowBackVowel)

In this formalism, elements to the left of a colon are lexical, while elements to the right are surface/realised, and the  $\Leftrightarrow$  denotes bi-implication. The rules state that a lexical high harmonising vowel is realised as round if and only if it follows a surface round vowel (and any intervening consonants), and that a high (resp. low) harmonising vowel is realised as back if and only if it follows a surface back vowel (and any intervening consonants). The formalism is straightforward to understand, and more importantly, easily implementable as a collection of finite-state transducers that can

subsequently be cascaded or intersected into a single transducer, allowing for efficient evaluation of phonological systems (Beesley and Karttunen, 2003).

### 2.4.3 LANGUAGE CHANGE MODELLING

Computational and mathematical modelling of language change is a relatively recent development in linguistics, with the bulk of the research in this domain having been carried out in the last decade, and mainstream acceptance only in the last few years (Klein (1966); Klein et al. (1969) being rather startlingly prescient exceptions). The work done to date has mostly dealt with syntactic (Niyogi and Berwick, 1998) or morphological (Hare and Elman, 1995) change, with little work on phonological change until quite recently (Dras and Harrison, 2002; Wedel, 2007; Choudhury, 2007).

#### KIRBY (1996, ET SEQ)

In his PhD dissertation and subsequent work on language evolution, Kirby (1996, 1999) introduced *iterated learning* models, discussed above in subsection 2.3.2. IL models are essentially discrete dynamical systems that look at how data distributions change over time, when the output data from one iteration of learning (irrespective of the learning algorithm used) become the input data for the next learning iteration. Kirby introduced this

model to explicitly capture the feedback loop between *I*- and *E*-language in language transmission and acquisition. More explicitly, the learners of generation  $n + 1$  create their *I*-languages on the basis of the *E*-language of generation  $n$ .

Kirby showed that a rudimentary form of syntactic compositionality could arise from an initially non-compositional state, and argued that this is a result of the “information bottleneck” that occurs in acquisition, *i.e.* the fact that learners are exposed to a small subset of the possible input data forces them to generalise, and forms that are easy to generalise over, *e.g.* compositional forms, are more likely to be successfully transmitted. The general framework of iterated learning has proved popular in subsequent language modelling and been taken up by several authors (*cf.* Kirby et al., 2007, for an overview).

#### **HARRISON ET AL (2002, 2003)**

Dras and Harrison (2002) and Harrison et al. (2003) create multi-agent simulations of the distribution of backness harmony over time in the Turkic languages, in particular aiming to model the “S-shaped” (logistic) trajectory that has been claimed to characterize historical language change. They give historical data showing full harmony in Old Turkic (*ca.* 8<sup>th</sup> c.), steadily declining in various descendant languages to the current levels of variation, from near-total harmony in Tuvan to absence of harmony in Uzbek.

They model a population of interacting Turkic speakers with horizontal information flow, initiated with 50% harmonic 1,000-word lexicons. At each interaction, an agent can choose with some probability—either fixed or proportional to the amount of harmony already in its lexicon—to harmonize or disharmonise a word that is transmitted to it. By manipulating the parameters of the model (specifically the likelihood of harmonisation), Harrison et al. are able to recreate logistic curves for the emergence of harmony. However, they fail to do so for its decay, hypothesising that the mechanisms of emergence and decay are different.

There are both conceptual and methodological aspects of this series of simulations that are problematic. The parameter which controls agents' decision to (dis)harmonise a word conflates several properties (coarticulation, lexical frequency, *etc.*), at least some of which should probably be investigated individually, and children directly inherit a subset of their parent's lexicon, crucially eliminating the interaction that I take to be key in this type of language change. Harrison et al. take it as given that a logistic trajectory of language change is correct for the case of Turkic harmony—although this is not obvious from the data they give—and adjust their parameters until they are able to produce one, an example of “data fitting” that goes against standard modelling (and scientific) practice. Finally, although the simulations involve stochastic parameters, the authors only show that S-shaped curves arises in particular instances, rather than as an average behaviour, again going against the emerging standard methodol-

ogy of stochastic computational modelling.

### **WEDEL (2004, ET SEQ)**

Wedel (2004b, 2006, 2007) uses exemplar-based models of synchronic knowledge coupled with an explicitly Darwinian evolutionary model to study the emergence of categorical patterns and regularity in phonology. Rather than studying specific phonological patterns, Wedel examines how selective pressures (derived from the production-perception feedback loop), coupled with the dynamics of lexically-biased exemplar models, lead to the emergence of (i) categoricity from initially gradient phenomena, (ii) contrast maintenance, and (iii) something akin to strict dominance in Optimality Theory.

### **CHOUDHURY ET AL. (2006)**

Choudhury et al. (2006) are concerned with creating computational models of real-world phonological change, specifically the diachronic process whereby Sanskrit schwas were reduced to the point of deletion in modern Hindi. They build a synthetic simulation with horizontal information flow. Each agent broadly follows the architecture in (Russell and Norvig, 1995) and is structured as follows:

- Grammar: lexical entries as sequences of <phoneme, underlying

duration> pairs

- Articulator: <phoneme, surface duration> pairs and <phoneme-phoneme transitions, duration> pairs, coupled with inherent stochastic bias to reduce duration of schwa
- Perceptor: the string perceived is the maximum-likelihood parse of sequence of phonemes, given duration and transitions; the perceived word is closest word in lexicon based on edit distance

Choudhury et al. embed their agents in a *language game* scenario, where randomly selected pairs of agents interact and explicitly signal successful communication. Using an artificially normalised lexicon, they successfully model the reduction of schwa in positions roughly corresponding to those in Hindi.

Although it is a common framework for agent interaction in synthetic language modelling (Steels, 1997; de Boer, 2001), it is clear that the mode of feedback in language game scenarios is unrealistic, especially since the feedback is about structure. It is generally accepted that feedback, especially in language acquisition, does not concern the formal properties of language (Marcus 1993, but see Chouinard and Clark 2003 for a more nuanced view). More problematic in these simulations is that the built-in stochastic pressure for context-free schwa reduction seems to build the looked-for behaviour right into the model. One wonders what else but the emergence of schwa deletion was expected, given a steady stochastic bias

towards the shortening of schwa.



## CHAPTER 3

# ACQUISITION

The last decade has seen a move away from rationalist theories toward empiricist methods in phonology.<sup>1</sup> One of the questions driving this shift is: *How much of the technical baggage of our theories can be inferred from data, i.e. other than by positing domain-specific cognitive biases?* Examples of recent work in this vein are Mielke's (2008) work on emergent features, and recent probabilistic and information-theoretic approaches to learning phonological categories and processes (*cf.* Goldwater and Johnson, 2004; Goldsmith and Xanthos, 2009). As discussed in subsection 2.3.3, among empirically-oriented models, we can draw a distinction between those that use *eager* versus *lazy* methods (Aha, 1997). The vast bulk of the empiricist

---

<sup>1</sup>There are in fact two major distinctions that cross-classify phonological theories: rationalist (*i.e.* nativist) vs. empiricist, and symbolic vs. distributed/sub-symbolic. The near-absence of sub-symbolic, rationalist approaches is a contingent fact. Such theories are, in fact, possible (*cf.* Marcus, 2001, for an example).

research carried out in phonological modelling has implicitly (in the case of theoretical work) or explicitly (in the case of implemented models) made use of eager methods and symbolic representations *modulo* some work in connectionist phonology, *e.g.* Gasser and Lee (1992); Joanisse (2000)). One of the goals of this chapter is to show that a lazy algorithm, in concert with unstructured numerical representations, can account for the acquisition and use of a generalised, productive synchronic pattern.

In this chapter I present a lazy learning model of phonological acquisition and usage, `LIBPHON`, which learns a pattern of synchronic alternations that closely mimics the morphophonological phenomena that characterise progressive stem-controlled vowel harmony, including learning about opaque and transparent vowels. I begin in section 3.1 with a high-level view of `LIBPHON`'s basic data structures, and perception and production algorithms, paying particular attention to the mechanism by which it analogises over memorised tokens. Section 3.2 details the task that `LIBPHON` faces. Then, in section 3.3 and section 3.4, I present the simulations run on various parameterisations of `LIBPHON` and their results. I conclude with some discussion in section 3.5.

## 3.1 LIBPHON

LIBPHON, the **L**azy **I**nstance-based **P**honologist, is an artificial agent designed to model the core aspects of the synchronic acquisition and productive use of vowel harmony (and by hypothesis, any phonetically natural phonological assimilation; *cf.* section 1.1).<sup>2</sup>

Russell and Norvig (1995) propose that the defining properties of an intelligent agent, whether artificial or natural, are its means of interacting with the world in which it is embedded, *viz.* its sensors and effectors, and its internal state. The high-level architecture of my agent, shown in Figure 3.1, is heavily inspired by this view.

In lieu of sensors and effectors that enable interaction with a world, LIBPHON is endowed with speech production and comprehension modules, which enable linguistic interaction with other agents. Its internal model consists of a lexicon of stored linguistic experiences, coupled with an algorithm for analogising over subsets of the lexicon, permitting the comprehension and production of novel inputs and outputs. As they interact, agents can update their internal model by acquiring new expe-

---

<sup>2</sup>The models were programmed in Python, making heavy use of the numerical and scientific packages NumPy and SciPy (Oliphant et al., 2001). Python’s syntax lends itself to rapid development and easy legibility, and the data-intensive portions of NumPy and SciPy are implemented in C and FORTRAN, ensuring the speed of highly optimised compiled code.

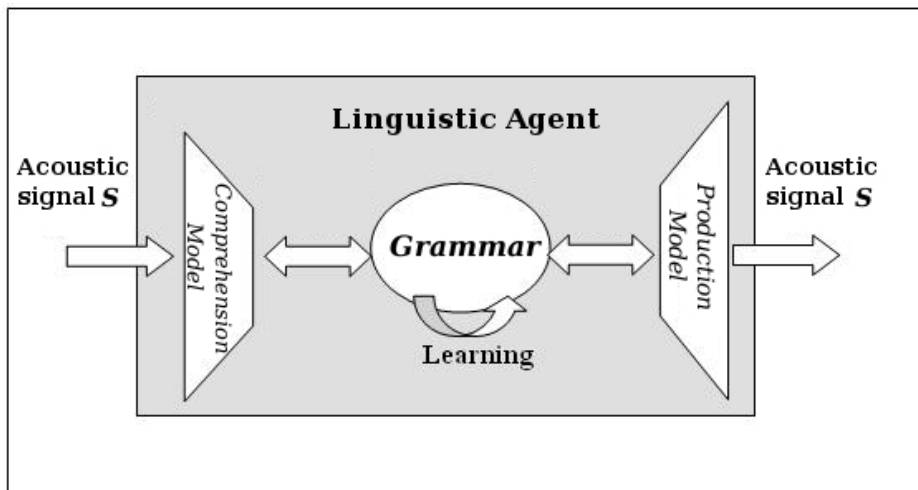


Figure 3.1: Schematic view of linguistic agent

rential tokens and making generalisations over the tokens they already have stored.

### 3.1.1 *Decisions & mechanisms REDUX*

Johnson (2007) presents a set of desiderata to be addressed in the development and implementation of an instance-based model of phonological knowledge: the decisions to be made concern the size and type of the basic units of analysis, and the mechanisms for similarity-matching and activation spread in memory.

## UNITS

**SIZE** One of the first issues faced in designing an instance-based model of cognition concerns the “granularity of experience”, *i.e.* the size of the perceptual/experiential tokens which are to be stored in memory. The principal distinction made in defining the size of the basic units in an instance-based model of phonology is *lexical* versus *sublexical*. These terms respectively cover any domains larger or smaller than a word, *e.g.* multi-word collocations, syllables, segments, *etc.* Both sizes of units, either separately or together, have featured in previous instance-based models of phonetics or phonology (*cf.* the segment/phone-sized units used in Pierrehumbert’s (2001) model, Silverman’s (2006) word-sized chunks, and Wedel’s 2004b use of both, *inter alia*).

I take meaning-bearing unanalysed acoustic “chunks”—words, broadly construed—to be the correct basic unit of analysis in instance-based language models, and hence in LIBPHON. The assumption that some degree of word-level segmentation of the speech signal is available to the language learner prior to acquisition of phonological phenomena is fairly uncontroversial, given evidence from acquisition studies (Ferguson and Farwell, 1975; Jusczyk, 1999). Motivation for the word-based view of exemplar storage comes from several sources. Johnson (2007) and Välimaa-Blum (2009) suggest that stored perceptual tokens are “of experiences”, and that what we experience in speech are word-sized chunks rather than *e.g.* features

or segments, citing as evidence the word-boundedness of lay-people's discussion of linguistic phenomena and properties. Johnson (1997a) also cites claims from *auditory scene analysis* (Bregman, 1990) that auditory streams have beginnings and ends, but no internal structure. Perhaps the strongest argument for a word-based primitive unit of storage is the wealth of empirical evidence (thoroughly reviewed in Lodge, 2009) that phonemic, or at least segmental, awareness is a direct result of alphabetic literacy training, *i.e.* that illiterates and literates of non-alphabetic languages have very poor phonemic awareness, with respect to monitoring, segmentation, and manipulation (Morais et al., 1979; Read et al., 1986; Faber, 1992; Carroll, 2004; Port, 2007a).

A consequence of the word-based approach adopted here is that generalisations about the behaviour of the learner (during or after acquisition) that can be described in terms of sublexical units (*e.g.* features, segments, *etc.*) must be viewed as *emergent* properties either of the dynamics of learning over whole-word sequences, or of analogical computation over a lexicon.

**TYPE** Having determined the appropriate “size” of LIBPHON's basic unit, I turn now to the issue of deciding on the *type* of the space over which these units are defined, *i.e.* whether their components are continuous or discrete, gradient or categorical. To begin with, it must be possible to define a similarity function in this space of features, since LIBPHON is an instance-based model, and its classification/generation algorithms require explicit

similarity comparisons. The similarity function that I use here is based on Euclidean distance, and so units must be embedded in a numerical metric space. The dimensionality of the chosen space can drastically affect the efficiency/tractability of a particular implementation; naive implementations of instance-based learning result in query-time that is exponential in the dimensionality of the instance-space (Aha et al., 1991). In low-dimensional spaces like the one I use here, this is not a problem, but it can be computationally prohibitive when the dimensionality is high. Given that LIBPHON is meant to model phonological phenomena, an obvious candidate is traditional phonological features.

Since the middle of the 20<sup>th</sup> century (*cf.* Chomsky and Halle, 1968), theories of phonology have nearly all made the (sometimes implicit) claim that lexical representations are stored in memory in terms of features that either have articulatory phonetic content, or that transduce directly to articulatory instructions (*cf.* Halle, 1997, for explicit discussion of this viewpoint). Coleman (1998, 2002) and Phillips (2001), citing evidence from the neuroscientific and psycholinguistic literature on lexical representation (*e.g.* from speech perception and phoneme monitoring experiments), show that evidence for this position is weak at best, and that lexical representations are more likely to be acoustic than articulatory. Hickok (2009), reviewing neuroscientific literature on the role mirror neurons play in action understanding, finds no clear contribution of motoric (hence, articulatory) sys-

tems in speech comprehension.<sup>3</sup> Finally, Mielke and Lin (2007) and Mielke (2008) provide evidence for and models of distinctive feature learning. But if features are learned, it must be on the basis of input, and the only input immediately accessible to the learner (*pace* Fowler 1986) is acoustic.

In light of the preceding arguments, I take the type of LIBPHON's instance space to be real-valued and ranging over acoustic features. In particular, I use acoustic formants—bands of increased energy in the spectral representation of human speech (Johnson, 2003)—as the embedding dimension for the simulations discussed here (see Johnson, 1997a; Kirchner and Moore, 2009, for other approaches), as it has been shown that speech tokens synthesised purely from formant information can reliably be recognised by humans (Cooper et al., 1952), and the low dimensionality of this representation makes the simulations more computationally tractable. Vowels are specified by their midpoint formant values and consonants are specified by so-called “locus” values, which can be identified by inspecting the trajectories of consonant-vowel transitions in speech (Lindblom, 1963; Sussman et al., 1998).

Since the particular phenomenon I am addressing is palatal harmony, and  $F2$  value is a reliable acoustic correlate of vowel backness (Johnson, 2003), LIBPHON's representations are restricted to  $(F1, F2)$  trajectories. Hence the

---

<sup>3</sup>Evidently, articulatory systems *are* relevant in the actual production of speech. I have abstracted away from this influence in this thesis, but there is good reason to believe that articulatory/auditory feedback in self-generated forms plays an important role in mastering at least some aspects of phonology (Juszyk, 1999).



basic unit of analysis in LIBPHON is a word-sized sequence of  $(F1, F2)$  values, which I will call a *trajectory*.

#### A NOTE ON TIME

For the simulations described here, I have chosen to use fixed-rate trajectories, in which consonants and vowels are represented in a temporally coarse-grained manner with single  $(F1, F2)$  tuples. Of course, consonants and vowels in actual human speech unfold in time, and are more realistically represented as formant sequences themselves. Modelling trajectories at this level introduces the problem of temporal variability; repeated tokens of a given word—both within and across speakers—vary widely in duration. This variability is one of the biggest obstacles in the development of instance-based models of speech production, due to the difficulty of aligning variable-length forms.

The algorithms that exist for aligning variable-length real-valued sequences typically involve sophisticated dynamic programming algorithms (e.g. dynamic time warping and hidden Markov models Rabiner and Juang, 1993), the details of which are not particularly cognitively plausible. Kirchner and Moore (2009) use dynamic time warping in their model, PEBLS (2.2.2); an instance-based production-oriented model of spirantisation using real (*viz.* variable-rate) speech signals.

**LANDMARKS** It is clear that information about speech rate must be stored somehow; we are able to identify tokens of speech as relatively fast or slow, we can have a memory of some interlocutor as a particularly fast or slow speaker, and we can voluntarily imitate fast or slow speech. A possible justification for my choice of fixed-rate representations would be to claim that rate information about tokens of speech is not stored in the same way that acoustic information is (perhaps in the same way that we store *e.g.* indexical information about speaker identity). On this view, rate information is somehow “parsed away” from the acoustic information. A good candidate for how this could happen is (a modified version of) Stevens’s (2000; 2002) landmark-based account of lexical access. On this view, our perceptual systems are sensitive to “acoustic landmarks” in the speech signal (*e.g.* peaks and troughs in frequency-specific amplitudes or spectral discontinuities). Whenever a landmark is detected we “poll” our systems for the values of some relevant acoustic parameters (here, formant values) and it is atemporal (but ordered) sequences of these values that are stored. This is a highly simplified account which (i) departs non-trivially from Stevens’s view, and (ii) has many holes that would have to be patched. My motivation for introducing it here is simply to show that a fixed-rate representation is at least plausible.

**REVISITING COARTICULATION AND PERCEPTUAL COMPENSATION**

**COARTICULATION** As I hinted previously, the model of phonological knowledge being developed here as *LIBPHON*, according to which humans have an instance-based *I*-language and acoustically-represented lexical entries, renders the issue of within-word coarticulation largely moot. Synchronically, the fact that I pronounce *keep* as [k<sup>j</sup>ip] and *coop* as [k<sup>y</sup>up] is a consequence of the fact that this is how I have heard them pronounced. Note, however, that there is still a role for coarticulation due to speech planning in producing novel or rare multi-word sequences and root-affix combinations, although it is likely the case that many frequent multi-word sequences are also stored as units in our lexicons (see subsection 2.1.4).

**PERCEPTUAL COMPENSATION** In an acoustic word-based exemplar model, an “active” mechanism of coarticulatory compensation is unnecessary, since similarity judgments are made over entire acoustic trajectories, and the variability attributable to coarticulation under a traditional model is typically not severe enough to jeopardize them. In cases where this is false (*e.g.* severely degraded inputs due to environmental noise), one might invoke independent pressures against homophony. Finally, it seems likely that the demonstration of perceptual compensation in (Beddor et al., 2002) could be at least partially attributed to task effects: the task that the participants are faced with, same-different judgments on cross-spliced syllables,

has dubious ecological validity at best. It is similar in nature to phoneme monitoring, which has been shown to occur *after* lexical access (Démonet et al., 2002). Moreover, the task explicitly requires the interpretation of novel, nonreal junctures, hence the juncture effects mentioned in a previous quote by Coleman may be at play.

## 3.2 THE TASK

In training/acquisition, LIBPHON is given input in the form of paired category labels—“meanings”—and tokens of their realisations, which I denote by (LABEL, instance) tuples. The structure and content of both LABELS and instances are discussed below. LIBPHON’s success in acquiring some generalised pattern is tested by either (i) giving it an unlabelled instance and asking for an appropriate LABEL, or (ii) giving a previously unseen class LABEL for which it must generate a representative instance.

### 3.2.1 THE LANGUAGES

In order to get at the essence of the problem under consideration, *viz.* the acquisition of vowel harmony as traditionally characterised by morphophonological alternations, LIBPHON learns simplified artificial languages which have just enough structure to capture the phenomena of interest, rather than being trained on real-world harmonic data. This is

also in the interest of computational tractability, and obviates the need for data collection in languages that are not readily available. As discussed in section 2.2, in an instance-based approach, phonological knowledge is taken to emerge from generalisations over lexical items, and so the key to acquiring any given phonological pattern lies in learning a lexicon that instantiates that pattern. This view of phonological learning is relatively uncontroversial in the acquisition literature (*cf.* Ferguson and Farwell, 1975; Jusczyk, 2000, and references therein). Consequently, the languages learned by LIBPHON abstract away from sentence-level phenomena—although the acquired units could in principle be “undersegmented” multi-word collocations—and the linguistic input simply takes the form of word-sized formant trajectories.

**PHONOLOGICAL INVENTORY** The phonological inventory for the simulations in this chapter consists of three consonants and four vowels, which I will label {b, d, g} and {i, e, u, o}, respectively. Recall that LIBPHON’s basic units are defined over a parametric acoustic space, and so these labels have no theoretical status, and are simply for mnemonic convenience. Vowel formant values were generated from formant synthesis equations in de Boer (2001) and consonant values were derived from locus equations in Sussman et al. (1998) (see Appendix A). The {i, e} and {u, o} vowels have high  $F_2$  and low  $F_2$ , respectively.

**LEXICAL ITEMS** As discussed above, instances of formant trajectories are associated to category labels, which from a formal point of view are contentless indices, but are meant to model both arbitrary semantic content and some grammatical properties of (traditionally construed) morphemes. These are compositional, comprising:

- an obligatory “lexical” meaning, denoted by a quasi-arbitrary CVCVCVCV string, corresponding (very) loosely to its associated trajectories’ English pronunciation,
- one of two obligatorily present “case markers”, NOM and ACC, and
- and an optionally present “plural marker”, denoted PL.

I refer to these category labels generically as LABELS. The terms *case* and *plural* are used mnemonically here, as the languages modelled have neither syntax nor semantics, but these markers are meant to bear some relation to properties of natural languages. Hence, LABELS in the artificial languages come in four forms, LEX NOM, LEX PL NOM, LEX ACC, and LEX PL ACC (there is no category label corresponding to singular).

Figure 3.2 gives an example lexical entry for a “singular nominative” form. The nominative is the default form, with zero exponence. The lexical meaning, GIDEGEBI, is associated with a set of 2-dimensional formant trajectories, (labelled *F1* and *F2* in the figure for expository convenience, although these are not explicitly represented in the implementation). I show some

GIDEGEBI NOM	$\left[ \begin{array}{cc} F1 & F2 \\ 121 & 2096 \\ 391 & 2118 \\ 122 & 2251 \\ 248 & 2096 \\ 115 & 1821 \\ 405 & 2015 \\ 116 & 2331 \\ 398 & 1993 \end{array} \right]$	$\left[ \begin{array}{cc} F1 & F2 \\ 130 & 2104 \\ 388 & 2111 \\ 123 & 2260 \\ 250 & 2075 \\ 111 & 1833 \\ 408 & 2008 \\ 122 & 2327 \\ 400 & 2001 \end{array} \right]$	$\left[ \begin{array}{cc} F1 & F2 \\ 125 & 2099 \\ 378 & 2113 \\ 125 & 2254 \\ 247 & 2083 \\ 113 & 1828 \\ 410 & 2017 \\ 121 & 2333 \\ 398 & 1997 \end{array} \right]$	...
--------------	--	--	--	-----

Figure 3.2: Example of lexical entry for GIDEGEBI NOM

graphs of lexical forms below which perhaps render the representation more perspicuously.

**HARMONISING SUFFIXES**

For many generative phonologists, lexical generalisations, while perhaps indicative, are not truly diagnostic of vowel harmony, which is taken to be characterised by a productive and general synchronic system of alternations (*cf.* the quote by Mahanta in subsection 2.1.3). In typical cases, we find affix allomorphs which are realised with alternating vowels, depending on the vowels in the stem, as illustrated in subsection 2.1.1.

Hence, the simulations described here featured an alternating set of affixes, in addition to the lexical harmony described above. The `NOM` case marker has null realisation, whereas the `ACC` marker surfaces as either [be] or [bo],

with the  $F2$  value of the vowel determined by the root vowels.

### NEUTRAL VOWELS

The harmony processes examined thus far are in some sense “local”, being describable in terms of adjacency, *e.g.* on a hypothesised autosegmental vowel tier (although the presence of intervening consonants still renders the harmony process “nonlocal” in some more concrete articulatory sense, *pace* Gafos 1999). One of the hallmarks of vowel harmony, as discussed in subsection 2.1.2, is the phenomenon of *neutral vowels*. These vowels fail to undergo harmony, and may or may not induce harmonic alternations in other vowels within the relevant domain. To introduce a neutral vowel, I added a category label,  $PL$ , whose realisation corresponds roughly to [gu], and which is treated as being either opaque or transparent in the simulations described below.

Each of Figure 3.3, Figure 3.4, and Figure 3.5 gives a graphical representation of the various “inflected” forms of the lexical meaning GIDEGEBI. The even-numbered indices on the  $x$ -axis correspond to consonants and the odd-numbered indices, the “pinches” in the graph, to vowels. Figure 3.4 and Figure 3.5 illustrate the difference between languages with opaque versus transparent  $PL$ , as reflected in the realisation of the ACC marker, which agrees in  $F2$  with the realised form of the  $PL$  exponent or of the root, respectively.



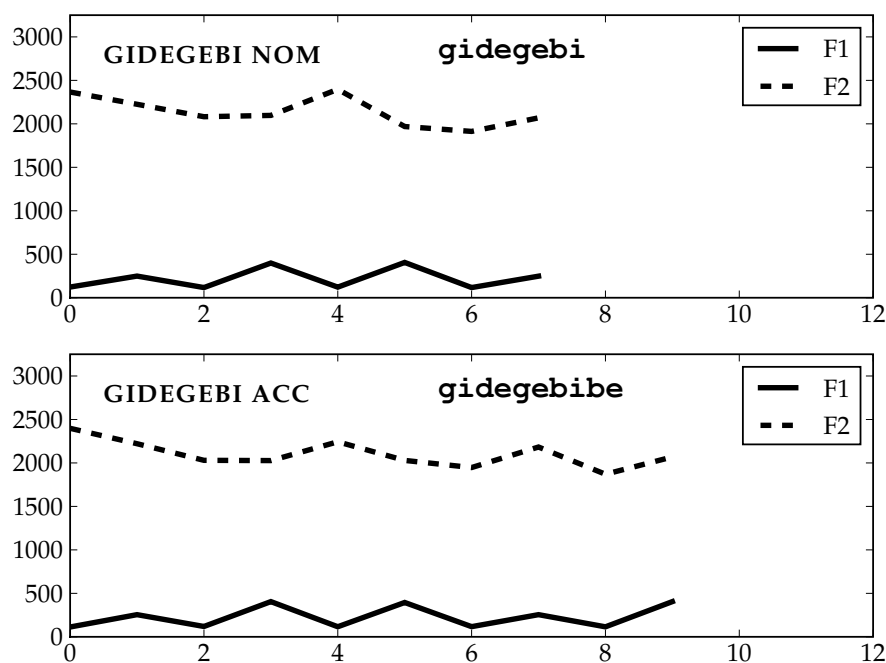


Figure 3.3: Graphical representation of singular forms of GIDEGEBI, as produced by teacher agent

Table 3.1 and Table 3.2 give illustrative data for the language cases in which the PL realisation is opaque or transparent, respectively. Each column gives a sort of mini-paradigm of the surface realisations of the various forms of a given lexical meaning, one with high *F2* realisation, the other low (with traditional morphological boundaries indicated for expository purposes). These languages capture the core properties of Tangale and Wolof, seen above in subsection 2.1.2.

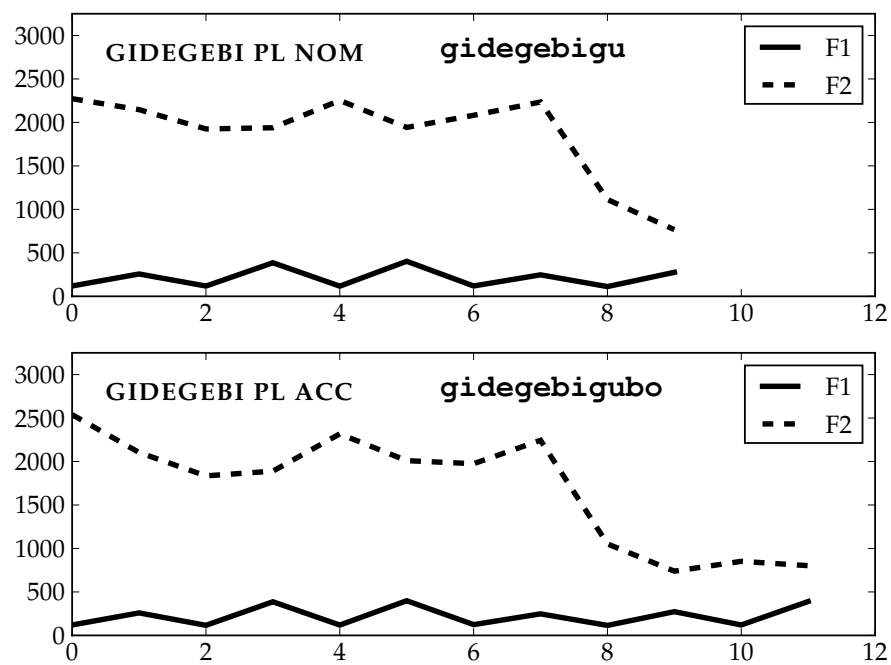


Figure 3.4: Graphical representation of plural forms of GIDEGEBI, as produced by teacher agent, with opaque PL realisation.

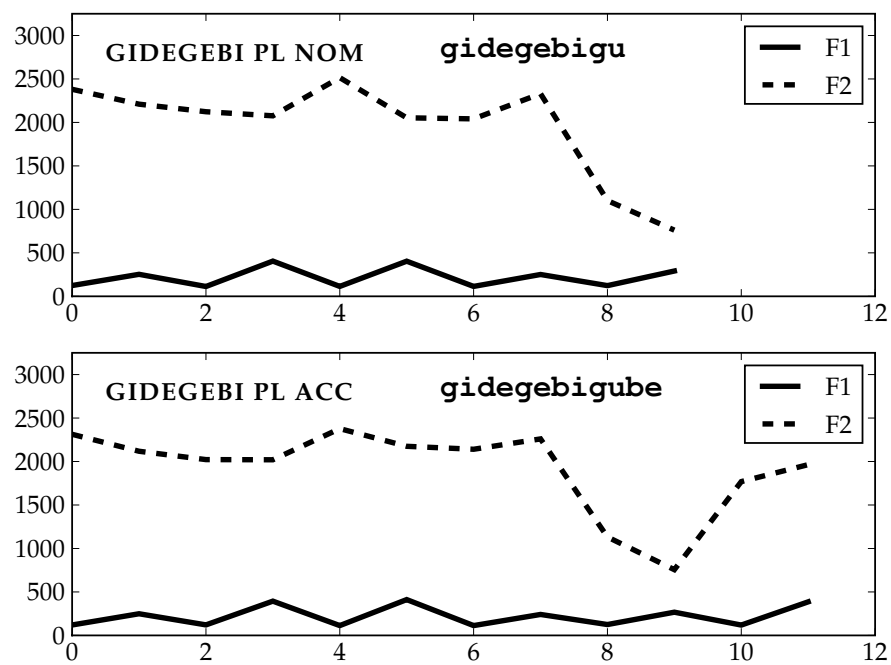


Figure 3.5: Graphical representation of plural forms of GIDEGEBI, as produced by teacher agent, with transparent PL realisation.

	High <i>F2</i> root	Low <i>F2</i> root
NOM	gidegebi	gudogobu
ACC	gidegebi-be	gudogobu-bo
PL-NOM	gidegebi-gu	gudogobu-gu
PL-ACC	gidegebi-gu- <b>bo</b>	gudogobu-gu- <b>bo</b>

Table 3.1: Opaque neutrality in LIBPHON’s languages

	High <i>F2</i> root	Low <i>F2</i> root
NOM	gidegebi	gudogobu
ACC	gidegebi-be	gudogobu-bo
PL-NOM	gidegebi-gu	gudogobu-gu
PL-ACC	gidegebi-gu- <b>be</b>	gudogobu-gu- <b>bo</b>

Table 3.2: Transparent neutrality in LIBPHON’s languages

The final row of the “High *F2*” column in each of the tables illustrates the variety of neutrality of the PL realisation. In Table 3.1, the opaque case, the realised form of ACC agrees in *F2* value with the preceding PL realisation, whereas in the singular form it agrees with the *F2* of the root vowels. In Table 3.2, the transparent case, the realisation of ACC has high *F2*, in agreement with the root vowels, notwithstanding the presence of the intervening low *F2* PL realisation.

### 3.2.2 PERCEPTION/CATEGORISATION

LIBPHON’s method of perception/categorisation of inputs is a relatively standard nearest-neighbour-based classification algorithm. See Algo-

rithm 1 for a pseudocode description.

---

**Algorithm 1** PERCEIVE(*input*, *k*)

---

**Require:** *input* as (LABEL  $\in$  [LEX](PL)[NOM | ACC], *instance*  $\in \mathbb{Z}_{2 \times 8, 10, 12}$ ), *k*  $\in \mathbb{Z}$

```

if LABEL is not empty then
  if LABEL  $\notin$  lexicon then
    Create LABEL in lexicon
  end if
else
  neighbours  $\leftarrow$  k-nearest neighbours of instance
  LABEL  $\leftarrow$  majority class label of neighbours
end if
Associate(instance, LABEL)

```

---

Inputs are (LABEL, *instance*) pairs, where LABEL can be empty. If LABEL is not empty, LIBPHON checks its lexicon to see whether it has any instances associated to that LABEL. If so, it adds the input to the set of trajectories associated with the input label. If it has no matching label, a new class LABEL is created and *instance* is added as its sole associated trajectory. If LABEL is empty, LIBPHON assigns *instance* to the majority class of its *k* nearest neighbours.

### 3.2.3 PRODUCTION/GENERATION

In production, LIBPHON is given a LABEL, and has to generate an appropriate trajectory for it. Recall that labels are compositional, with accessible subparts, and comprise (i) obligatory lexical meaning and case marker, and (ii) an optional plural marker. There are consequently several query types

to consider in generating output for a given LABEL.

In the two simplest cases, either the queried LABEL already exists in LIBPHON's lexicon, or else there are no LABELS with the same lexical sub-part (*i.e.* LIBPHON is being asked to generate a form of a word that it does not know). In the former case, a “seed” trajectory is uniform<sup>4</sup> randomly selected from the set associated with the queried LABEL, an analogical set is computed by finding the nearest neighbours to the seed within the associated set, and an output is generated by taking a distance-weighted mean over the analogical set. In the latter case, the query is simply ignored.

In the more interesting cases, LIBPHON has a LABEL in its lexicon with the same lexical meaning, but with differing CASE and/or PL specification. Consider the case in which LIBPHON knows only the singular NOM form of the queried label, but has to produce the PL ACC form. A seed instance is randomly selected from the set of trajectories associated to the NOM entry in the agent's lexicon, as this is the only entry with the corresponding lexical meaning, and it is a variant of this meaning that LIBPHON must produce. The analogical set, in this case, is composed of the seed's nearest neighbours in the set of all trajectories associated with LABELS of the form [LEX PL ACC], as LIBPHON is being asked to generate a PL ACC form. Once

---

<sup>4</sup>Exemplar models often bias the selection of seed tokens with “activations” by taking into account recency and frequency. Although the results presented below indicate that this is not necessary for ultimate attainment of the phenoma in question, it is likely that this kind of bias will need to be incorporated into LIBPHON to accurately model more nuanced effects.

again, the output produced is a distance-weighted mean of the analogical set.

This general procedure (*viz.* seed from a known item with same lexical meaning, analogical set from all items with desired inflection) is carried out in parallel cases with all other possible LABEL mismatches, *e.g.* a singular LABEL queried, but only a plural LABEL in the lexicon, a NOM query with only an ACC form in the lexicon, *etc.* In the cases where the lexicon contains multiple entries with the same lexical meaning, but not the query, the seed is selected from the LABEL with the closest semantic match. Algorithm 2 gives pseudocode for LIBPHON's production algorithm.

---

**Algorithm 2** PRODUCE(LABEL,  $k$ )

---

**Require:** LABEL  $\in$  [LEX](PL)[NOM | ACC],  $k \in \mathbb{Z}$

```

if LABEL  $\in$  lexicon then
    seed  $\leftarrow$  uniform random selection from instances associated to
    LABEL
    cloud  $\leftarrow$  all instances associated to LABEL
else if  $\exists$  LABEL'  $\in$  lexicon s.t. lex(LABEL') = lex(LABEL) then
    seed  $\leftarrow$  uniform random selection from instances associated to
    LABEL'
    cloud  $\leftarrow$  all instances associated to plural(LABEL)  $\cup$  case(LABEL)
else
    pass
end if

neighbours  $\leftarrow$   $k$ -nearest neighbours of seed in cloud
return distance-weighted mean of neighbours

```

---

**PRODUCTION AS REGRESSION**

The final steps in LIBPHON's production algorithm, finding the analogical set and computing the output as a weighted average, together constitute a technique known in the statistical learning literature as *kernel-smoothed nearest-neighbour regression*, and more specifically are closely related to the Nadaraya-Watson estimator (Hastie et al., 2009):

$$\hat{f}(x) = \frac{\sum_{i=1}^N K_\lambda(x, x_i) y_i}{\sum_{j=1}^N K_\lambda(x, x_j)}$$

with inverse-distance as the kernel-smoother,  $K$ , and the bandwidth function,  $h_\lambda(x)$  determined by the number of nearest neighbours,  $k$ . This link to the statistical learning literature puts LIBPHON on solid theoretical footing and opens the door to a variety of future research paths, *e.g.* experimenting with different kernels.

**A POTENTIAL CRITICISM**

There is a potential concern with averaging over unrelated forms (*i.e.* when the queried label is not in the lexicon) in order to generate an output: the forms returned may take intermediate values that are not representative of any particular LABEL's associated tokens (Janet Pierrehumbert, personal communication).



On the one hand, the fact that the averaging is weighted by distance militates somewhat against this problem: the relatively high likelihood that there are stored forms close to the seed that will disproportionately contribute to the generated output means that these situations will be rare. That being said, it is in fact possible for something like this to occur, and we shall discuss such a case in 3.4.2. The upshot is that *all* lazy learners have to have done some “learning” in advance of successful classification or generation (there can be no “nearest neighbour” in the absence of neighbours).

### 3.3 SIMULATIONS

The parameters of interest in assessing LIBPHON’s performance are: lexicon size, direction of information flow, and neighbourhood size used in production. The simulations here discuss the effects on learning of systematically varying each of these parameters.

#### 3.3.1 INFORMATION FLOW

The discussion in subsection 2.3.2 highlights the different possibilities for the flow of information in an agent-based simulation, in particular differentiating *vertical* and *horizontal* flows. One of the aims of the simulations carried out in this chapter is to assess the extent to which information flow

affects rate of acquisition. Below I discuss each of the possible flows, and its motivations, and present results of simulations instantiating those flows.

### (STRICTLY) VERTICAL

The simplest imaginable information flow in an agent-based model is what might be called *strictly vertical*. In this case, each learning agent communicates with exactly one teaching agent, as illustrated in Figure 3.6.

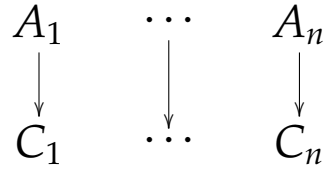


Figure 3.6: Strictly vertical information flow. Each child learns from exactly one adult.

Many iterated learning models in the literature take this form (cf. Kirby, 1999; Wedel, 2004b, among others), as it distills the acquisition scenario to its bare essentials. Clearly, of course, this is an unrealistic model, not least because comparatively few infants learn language from a single parent. This motivates the first generalisation of information flow.

**OBLIQUE**

In a simulation with *oblique* flow, learners get input from multiple adults, each of whom has individually variable outputs, as in Figure 3.7. On the one hand, this reflects the intuition that infants typically acquire language from a somewhat extended set of adults, including one or more parents, extended family, and other caregivers. In addition, this permits an initial foray into the exploration of the role that variation may play in acquisition.

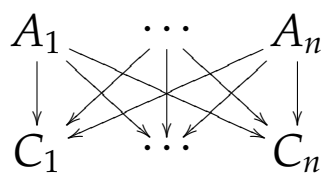


Figure 3.7: Oblique information flow. Each child learns from multiple adults, each with individually variable outputs.

It is by now standardly accepted in the literature on language acquisition that many of the detailed aspects of children's language learning (*e.g.* patterns of usage, intonation, etc.) are affected by interactions with other children, especially slightly older peers (Weinreich et al., 1968; Labov, 1994). The next logical step, then, is to introduce peer interaction into the acquisition model.

**VERTICAL/OBLIQUE WITH PEERS**

In each of these simulations, we model vertical and oblique information flow as above, but introduce horizontal, or peer, interactions as well, as illustrated in Figure 3.8 and Figure 3.9.

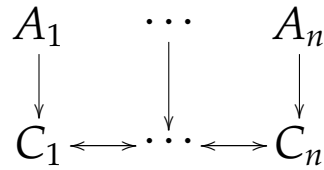


Figure 3.8: Vertical information flow with peer interactions. Each child learns from a single adult, plus other children.

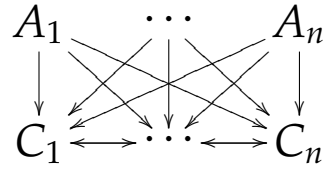


Figure 3.9: Oblique information flow with peer interactions. Each child learns from multiple adults and other children.

In order to model varying degrees of peer influence, I introduced a parameter,  $p$ , controlling the proportion of vertical versus horizontal information flow. On any given interaction, with probability  $p$ , the agent under consideration opts to learn from a (uniform randomly selected) peer, rather than

from a parent (in vertical flow) or other adult (in oblique flow).

## 3.4 EVALUATION

In this section I discuss the evaluative measures used for assessing LIBPHON's performance, and present the results of the simulations described above.

### 3.4.1 GENERALISATION

In computational learning research, as in linguistics, the notion of *generalisation* is important, in order to show that a learning algorithm has gone beyond simply memorising the data presented to it and succeeded in extracting some structure from the data. Assessing generalisation means assessing performance on novel data; the standard procedure for doing this in a machine/statistical learning scenario is to separate the available pool of data into *training* and *test* sets (Hastie et al., 2009). In all of the cases discussed here, LIBPHON was trained on a subset of the possible categories (typically on the order of 2500 out of approximately  $10^4$  possible LABELS). Test sets were created by selecting 500 LABELS that remained unseen at the end of training.

### 3.4.2 EVALUATIVE MEASURES

Assessing successful learning and generalisation in a computational model requires some measurable, quantitative outcome that can be tracked over time. Different learning tasks call for different evaluative measures, and it is important to choose measures that are informative, *viz.* that assess a learner’s performance on the task of interest. In particular, learners that classify data into a set of categories are typically assessed differently than learners that produce outputs. Although LIBPHON can do both of these things, my focus in this thesis is on generation and so I restrict discussion here to assessing LIBPHON’s outputs.

#### PRODUCTION/GENERATION

Although categorisation of novel inputs clearly plays a role in language acquisition, a phonologist studying acquisition is more likely to be interested in whether a learner has acquired a “productive rule”, *viz.* an input-output mapping that can generate correct outputs for novel or unseen inputs. This is essentially what the term “wug-test” (Ratner and Menn, 2000) has come to mean.<sup>5</sup> In the context of the current work, this corresponds to a scenario in which LIBPHON is given a previously unseen LABEL and must produce a

---

<sup>5</sup>In the original “wug test” (Berko Gleason, 1958), native English-speaking children aged 4-7 were shown drawings of novel animals and told their names (one of which was a “wug”), then plural forms were elicited to see if the children used the correct plural allomorph.

corresponding output.

Evaluative measures for this scenario typically involve a function of the error,  $E$ , in the learner's outputs, *i.e.* the difference between the output and some previously defined target. Common choices are functions of squared-error, as this is strictly non-negative and has a variety of mathematically-convenient properties (although these are typically only relevant in error-driven learning settings).

$$E = f((target - output)^2) \quad (3.1)$$

The measure of error I use is the per-LABEL root-mean-square deviation between LIBPHON's output and the teacher's mean output, averaged over the lexicon:

$$RMSE_{lex} = \frac{1}{|N|} \sum_{\hat{l} \in N} \sqrt{\frac{\sum_{i \in \hat{l}} (o_i - \hat{t}_i)^2}{n}} \quad (3.2)$$

In Equation 3.2, the innermost summation is taken pointwise over the difference between  $o$ , LIBPHON's output, and  $\hat{t}$ , the (pointwise) mean of the teacher's instances for some label in the test set,  $N$ . The outer terms sum this value over the entire test set, and then normalise to a per-test-item value. We are interested not so much in the magnitude of  $E_{lex}$ , as the manner in which it changes with growing  $|lex|$ .

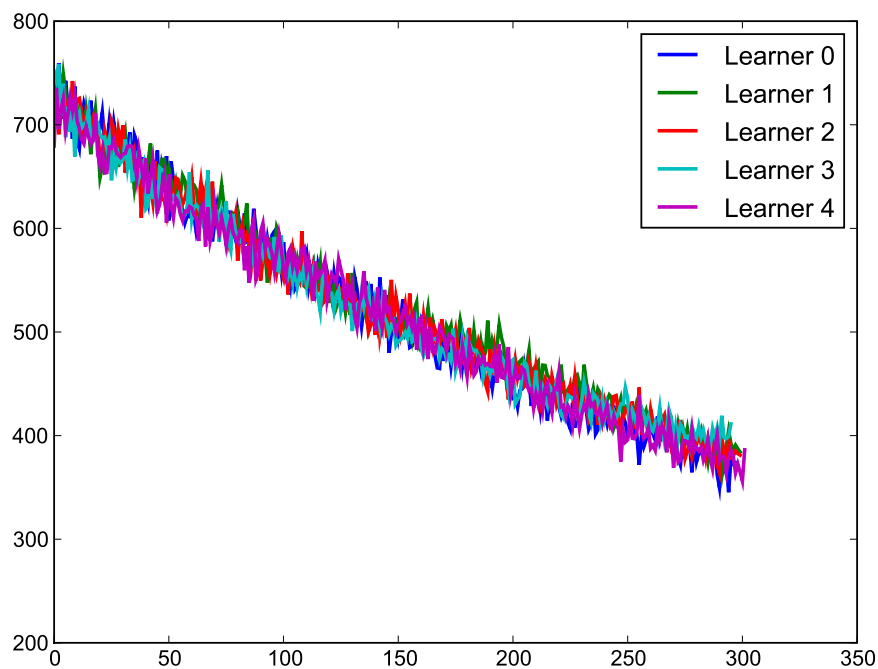


Figure 3.10: Root-mean-squared-error over test set as function of lexicon size. Vertical information flow, transparent neutrality.

Figure 3.10 and Figure 3.11 show the timecourse of the RMSE for vertical and oblique information flow, over the learning of 3000-word lexicons. Both qualitatively and quantitatively, the results are nondistinct, and in fact this is the case with the addition of peer interactions, as well, although there is slightly more variance in RMSE with peer interaction. Either nothing of interest is happening, or else the evaluative measure is obscuring important



information.

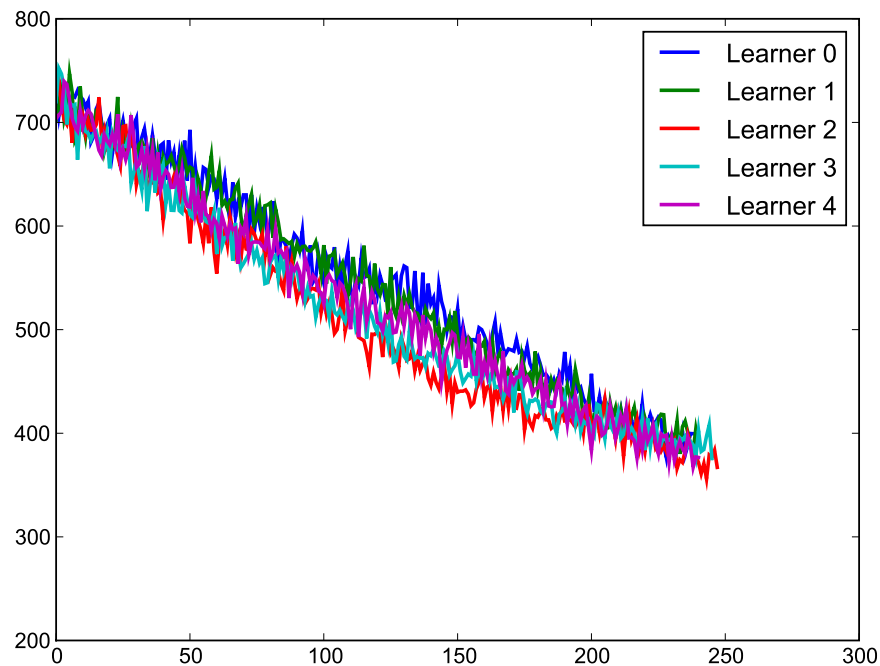


Figure 3.11: Root-mean-squared error over test set as function of lexicon size. Oblique information flow, opaque neutrality.

### IMPROVING ERROR MEASURES *VS* LEARNING VOWEL HARMONY

Although the evaluative measures discussed above are defined in such a way that their reduction is an indicator that LIBPHON is learning something, they do not provide a straightforward measure of the task that is really of interest, namely, the acquisition of a productive, generalised procedure

for generating correctly harmonised outputs. Consider the trajectories displayed graphically in Figure 3.12 and Figure 3.13 on pages 92 and 93. These show the evolution of LIBPHON's outputs for the various forms of GUBOGOBU in the relatively early stages of acquisition (up to 100 words), for a particular run of a simulation, with opaque PL realisation.

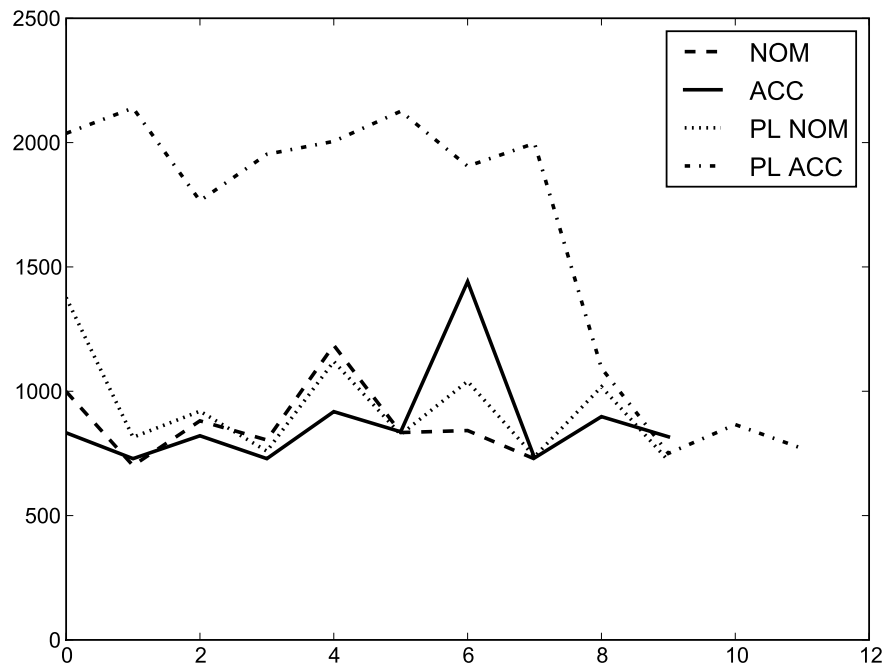


Figure 3.12: *F2* values for GUBOGOBU-\* forms. 10-word lexicon. Note outlier for PL ACC form.

There are two items of note here: (i) the initially outlying PL ACC form, and (ii) the speed with which it is rectified.

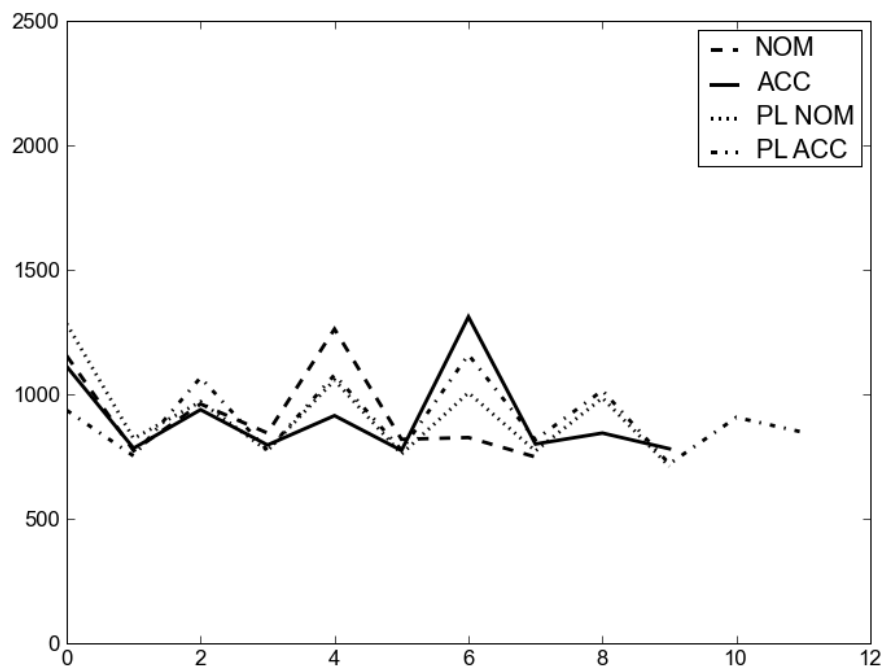


Figure 3.13:  $F_2$  values for GUBOGOBU-\* forms. By 30 words, some back PL ACC instances have been seen, outlier mostly regularised.

**EFFECTS OF SMALL LEXICONS** A first glance at Figure 3.12 immediately brings to mind the question *what's going on with the PL ACC form?* Fortunately, LIBPHON dumps its lexicon periodically during training, so we have the opportunity to investigate more deeply. Table 3.3 shows the contents of LIBPHON's lexicon at the 10-word stage in this run:

Of note, in particular, is the distribution of PL ACC, namely that by chance

<b>BIBEDIBE PL ACC</b>	GUGUBODU PL NOM
BUBUBODU ACC	DUBODODU NOM
<b>GEBIGIGI PL ACC</b>	DODODUGO PL NOM
GOBOGOGO NOM	GUBOGOBU NOM
DOBOGOBU PL NOM	<b>DIBIBEBE PL ACC</b>

Table 3.3: Contents of LIBPHON’s lexicon at the 10-word stage, for a particular run.

there are only high *F2* PL ACC forms at this particular point of LIBPHON’s acquisition in this particular simulation. Consider how the production algorithm works in the context of the present example. Given a label that is not in its lexicon, in this case GUBOGOBU PL ACC, LIBPHON checks whether it has any LABELS with the same lexical meaning, finding GUBOGOBU NOM in this case. A seed is then chosen from the trajectories associated to GUBOGOBU NOM, and the analogical set is computed from the seed’s nearest neighbours in *the set of instances associated to all PL ACC forms in the lexicon*, since LIBPHON is being asked to generate a PL ACC form. But, as mentioned above, these are all high *F2* forms in the present case, so the output form will necessarily have high *F2* (in the realisation of the root) as well.

By the time it has seen another 20 words in this run, LIBPHON has been exposed to some (in fact 2) low *F2* PL ACC forms, and is already well on the way to regularising its output. The output is still a bit peaky because the analogical set at this point still includes some high *F2* forms, although they are de-emphasised by the distance-weighted averaging process. The

take-home point here is that lazy classification or generation is impossible in advance of exposure to training data, and unlikely to be successful at the earliest stages of learning. In relation to human learning, this strikes me as an unproblematic property, as it is well-known that comprehension precedes production to a large degree in language acquisition, and it is almost trivially true that infants have been exposed to a good deal of input before they begin producing language.<sup>6</sup>

Now let us consider again the graphs in Figure 3.10 and Figure 3.11. There is simply no way to recover the information gleaned by looking at LIBPHON’s actual outputs on the basis of these kinds of measures, *i.e.* in assessing the acquisition of a productive, general rule of vowel harmony, this type of error measure is uninformative.

## 3.5 DISCUSSION

In this chapter, I presented LIBPHON, a model of the acquisition of vowel harmony (and by hypothesis all “phonetically natural” phonological assimilation). After motivating its general architecture and the particular set of decisions that went into its design (*viz.* instance-based learning and fixed-rate real-valued acoustic word-sized tokens), I showed that LIBPHON

---

<sup>6</sup>One hypothesis might be that children’s earliest productions, *e.g.* canonical and variegated babbling just *are* the outputs of an insufficiently-seeded lazy learning system (*cf.* Beckman and Edwards, 2000, for a similar view).

can learn a synchronic pattern of alternations that closely mimics the core features of vowel harmony, including patterns of opaque and transparent neutrality, and that the “rule” which is implicit in its lexicon and regression-based output procedure applies productively and generally to the classification and generation of novel forms. At minimum, LIBPHON provides an argument for the sufficiency of low-bias, domain general learning algorithms in the acquisition of some productive phonological patterns.

Standard evaluative measures indicate that LIBPHON’s performance improves as a function of lexicon size, but I also showed on the basis of LIBPHON’s outputs for a specific form that these measures are somewhat uninformative in assessing whether a productive harmony “rule” has been learned. This necessitated an examination of the evolution of individual forms, which revealed that harmony is learned on a case-by-case basis, sometimes quite rapidly. This is in accord with the few extant acquisition studies of vowel harmony, which indicate that infants achieve early mastery of harmonic forms, and rarely make mistakes (MacWhinney, 1978; Leiwo et al., 2006; Altan, 2007). The mistakes that do occur are on rare forms, exactly the situation in which LIBPHON is expected to err (recall the effect of sparsely-seeded regions of the lexicon).

## CHAPTER 4

# EVOLUTION

The simulations discussed in this chapter focus on a particular instance of historical explanation of a synchronic pattern: the diachronic evolution of vowel harmony. I will show that lexical harmony can emerge over generational timescales from interactions between synchronic coarticulation and a biased transmission-acquisition feedback loop. Depending on the amounts of coarticulation and channel noise, the amount of lexical harmony is seen to stabilize at intermediate levels between a baseline amount of harmony (given a uniform distribution over features) and full harmony.

## 4.1 MODELLING THE EMERGENCE OF HARMONY

### 4.1.1 THE AGENT

Recall that the general architecture of my linguistic agents follows that of Russell and Norvig (1995) (*cf.* Figure 4.1, repeated here from chapter 3), with comprehension and production modules in lieu of sensors and effectors, and a lexicon as internal state.

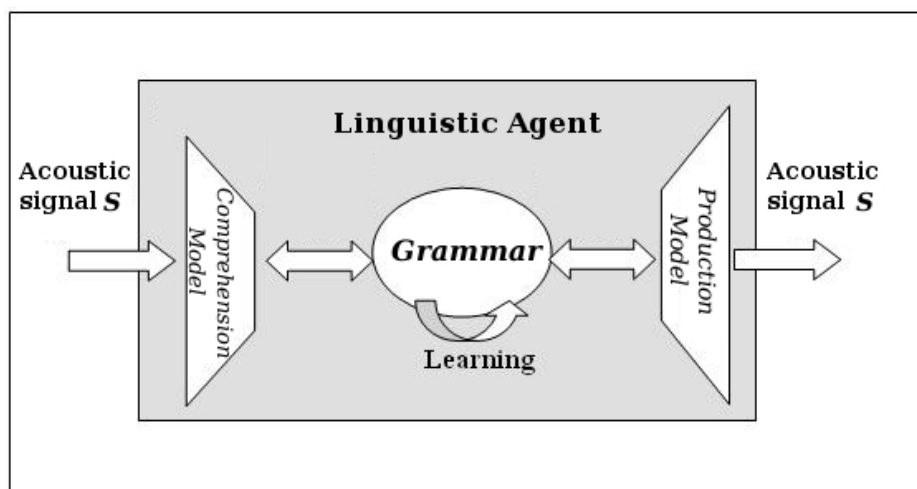


Figure 4.1: Architecture of a linguistic agent (adapted from Russell & Norvig 1995).

The agent in the simulations described in this chapter has a slightly more abstract lexicon than LIBPHON, being explicitly constructed from binary phonological features which model the traditional oppositions  $[\pm\text{COMPACT}]$  and  $[\pm\text{GRAVE}]$ .<sup>1</sup> In these simulations I abstract away from consonants

<sup>1</sup>I have opted to use Jakobsonian acoustic features rather than the more commonly-



altogether, and lexical items are sequences of four vowels, and there is moreover no morphophonology. This is obviously a highly impoverished “grammar”, yet it will be shown to suffice for the evolution of lexical harmony, given the learning algorithm discussed below.

In producing outputs, discrete phonological features are transduced to continuous articulatory parameters on the real interval  $[0, 1]$ , which are in turn used to synthesise  $F1$  and  $F2$  values using the same equations used to synthesise LIBPHON’s trajectories (see Appendix A). The articulatory specifications are Beta distributed (see Appendix B for details concerning the parameters) over the front/back space modelling individual-level hypo/hyperarticulation (Lindblom, 1990).

The parameter of interest in these models is variation in  $F2$ , which is the main acoustic correlate of front/back coarticulation. Both anticipatory and perseverative coarticulation were modelled by adding or subtracting a fixed amount from  $F2$  according to whether the preceding or following vowel had an opposite specification for GRAVE, and whether anticipatory or perseverative coarticulation is being modelled.<sup>2</sup> That is, in producing

---

known articulatory features of *SPE* because the parametric space I am interested in is acoustic, being defined over formants. The decidedly *nontraditional* use of binary oppositions on these features is for expository convenience; I will mostly be interested in polar shifts in  $F2$ . Whether these features are learned or innate is orthogonal to the discussion here, although the arguments in Mielke (2008) for emergence are persuasive. I assume their availability here for convenience.

<sup>2</sup>Of course, since I am varying purely acoustic representations, the term “coarticulation” is not, strictly speaking, appropriate, but this approach allows for more straightforward control of parametric variation, rather than manipulating the articulatory specifications that were fed to the synthesis equations described in Appendix A.

a  $V_1V_2$  vowel sequence such as *eo*, with anticipatory coarticulation, the  $F2$  value for *e* would be decreased by some amount,  $\delta$ , whereas with perseverative coarticulation, the  $F2$  value for *o* would be increased by  $\delta$ . Gaussian noise ( $\mu = 0$ ,  $\sigma^2 = 30\text{Hz}$ ) was added to the acoustic outputs to model the additive effects of extrinsic sources of noise such as articulatory fatigue, ambient acoustic interference, *etc.* This additional noise turned out to be of significance in the diachronic development of harmony.

The comprehension and learning modules are folded together in another multi-step system. The learner's inputs are length four ( $F1, F2$ ) trajectories, *i.e.* the adult's outputs. As an initial step, the learner uses  $k$ -means clustering (MacKay, 2002, p.285) to find acoustic prototypes in the data, with  $k$  set explicitly to 4. The simplification of essentially telling the learners how many vowels to look for was mainly in the interests of computational tractability, although it is also not implausible given the presumed availability of 2 binary features.<sup>3</sup> Once the acoustic prototypes (cluster means) have been found, the learner inverts the articulation-acoustics mapping to recover the articulatory parameters responsible for the data.<sup>4</sup> From the articulatory descriptions, the learner uses Bayes's Theorem to infer the underlying representations of each prototype vowel:

---

<sup>3</sup>Some attempts were made at learning the clusters with an EM-trained mixture of Gaussians pruned with the Bayesian Information Criterion, but the addition of coarticulatory effects renders the data non-Gaussian and the number of clusters is consistently overestimated.

<sup>4</sup>This is a somewhat implausible assumption, not least given that articulation supervenes on acoustics. I assume this can be addressed in future research, perhaps with an "analysis-by-synthesis" approach as advocated by Stevens and Halle (1967).

$$\hat{v} = \arg \max_v \frac{P(A = a|V = v)P(V = v)}{z}$$

Here,  $\hat{v}$  represents the learner's ultimate decision about the correct underlying form of the vowel under consideration,  $v$  is a hypothesis about the underlying form,  $a$  is the observed acoustic form,  $P(A = a|V = v)$  is the likelihood of the observed acoustic form, given the learner's hypothesis, and  $P(V = v)$  is the *a priori* probability of that hypothesis being correct ( $z$  is a normalizing constant to ensure that the calculation generates a probability distribution). This essentially corresponds to *Maximum A Posteriori* learning (Mitchell, 1997). Since I only investigate uniform priors (*i.e.* each underlying representation is equally probable, *a priori*), this algorithm reduces to Maximum Likelihood learning, whereby the underlying representation that gives highest likelihood to the observed acoustic form is the one chosen.

Given the articulatory specifications for the vowel cluster centres, the learner then assigns underlying representations to entire lexical entries by means of a simple vector quantization algorithm; each vowel in a word is assigned the underlying representation of the acoustic prototype nearest to it.

Algorithm 3 sketches the sequence of steps carried out for each generation of the iterated learning model incorporating the production and comprehension modules discussed above.

---

**Algorithm 3** Iterated learning algorithm for emergence of lexical vowel harmony

---

**Require:**  $gens \in \mathbb{Z}$  as number of iterations,  $coart_a, coart_p \in \mathbb{R}^+$  as anticipatory and perseverative coarticulation

Initialise zeroth ADULT agent with  $4^4$ -word lexicon (all length-4 permutations of 2 binary features)

```

while  $gens > 0$  do
  new CHILD (empty lexicon)
  ADULT outputs entire lexicon via formant synthesis, with addition
  of  $coart_a, coart_p$ 
  CHILD finds means of adult vowels via  $k$ -means clustering ( $k = 4$ )
  CHILD finds Max-Likelihood articulatory values for acoustic cluster
  centres
  CHILD assigns underlying reps to adult outputs from prototype ar-
  ticulations via vector-quantization
  ADULT  $\leftarrow$  CHILD (del previous ADULT)
   $gens \leftarrow gens - 1$ 
end while

```

---

### 4.1.2 SIMULATIONS

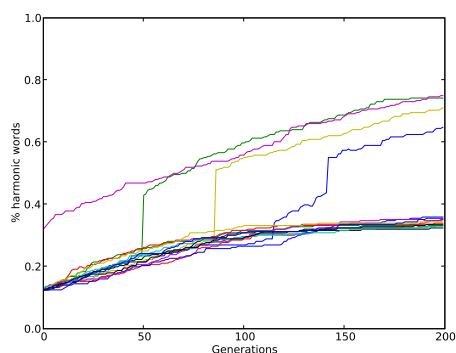
In the simulations described here, coarticulation was varied from 0 – –400Hz in 50Hz increments, a range which brackets the amounts of vowel-to-vowel coarticulation typically found in human languages (Öhman, 1966; Recasens, 1984; Magen, 1997; Manuel, 1999; Beddor et al., 2002). Each parameter setting was run on 15 agents for 250 iterations. The measure of interest in these simulations was the change in proportion of “fully harmonic” words in the learner’s lexicon—*i.e.* underlying forms with the same value for  $[\pm\text{GRAVE}]$  across all vowels—across generations. The graphs in

Figure 4.2 show the dynamics of the evolution of harmony for some of these parameter settings with anticipatory coarticulation. The results obtained for perseverative coarticulation were qualitatively and quantitatively similar and I do not discuss them further here.

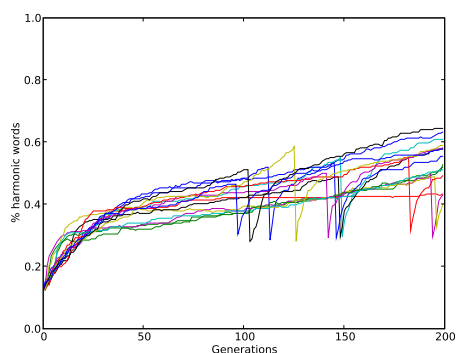
As is clear, the dynamics are non-trivial, appearing sometimes relatively monotonic, other times oscillating quite dramatically. Before discussing the specifics, it is important to note that the overall tendency *is* for harmony to increase across generational timescales, in some cases reaching full lexical harmony. This is an initial vindication of the Ohalian theory that I initially set out to assess in these simulations. Ohala's theory can be evaluated in modal terms, *viz.* as a statement about the necessity or sufficiency of vowel-to-vowel coarticulation for the eventual diachronic emergence of vowel harmony. The results here at least support sufficiency for the case under consideration: in the presence of vowel-to-vowel *F2* coarticulation that is perceivable by learners (and given that the lexical structure assumed here is not too far off the mark), vowels within lexical items will eventually come to agree with respect to the phonological feature  $[\pm\text{GRAVE}]$ .<sup>5</sup>

---

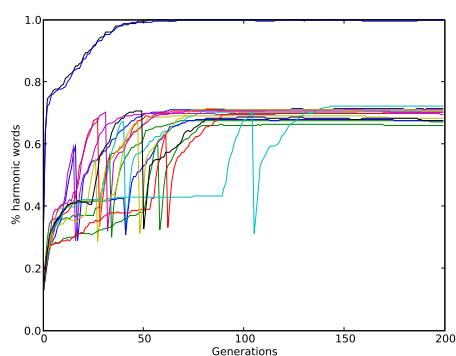
<sup>5</sup>This statement holds whether  $[\pm\text{GRAVE}]$  has some non-trivial cognitive status, or is simply a descriptive label for a set of physical/physiological facts.



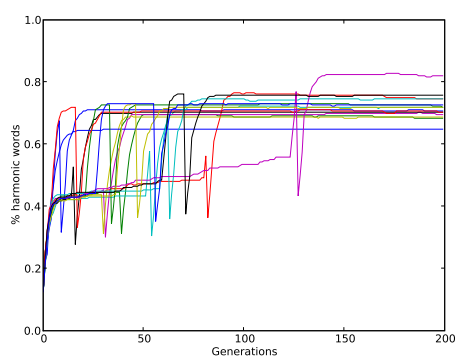
(a) 100Hz anticipatory coarticulation



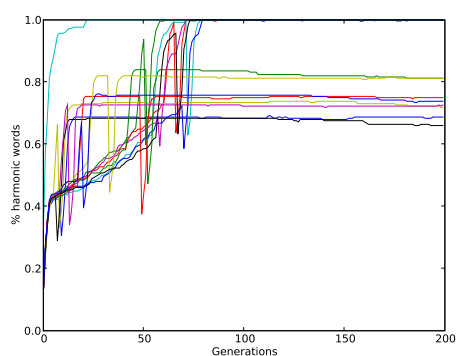
(b) 150Hz anticipatory coarticulation



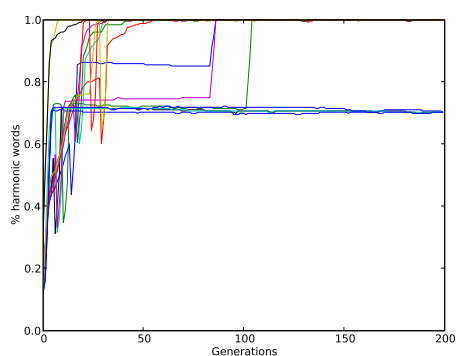
(c) 200Hz anticipatory coarticulation



(d) 250Hz anticipatory coarticulation



(e) 300Hz anticipatory coarticulation



(f) 350Hz anticipatory coarticulation

Figure 4.2: Effects of varying degrees of anticipatory coarticulation. 15 runs per figure. Gaussian noise  $\sim \mathcal{N}(0, 30)$  on outputs.

### 4.1.3 BIASED TOWARD SUCCESS?

In the face of steady stochastic pressure in the direction of harmonisation, all runs of the model might be expected to inexorably evolve toward fully harmonic forms across the lexicon (*cf.* the discussion of Choudhury's work in 2.4.3), and in some runs we do in fact see what looks like a relatively monotonic march toward harmony. This intuition is essentially a version of the "actuation problem" introduced by Weinreich et al. (1968), who note:

[...]the question always remains as to why the change was not actuated sooner, or why it was not *simultaneously actuated wherever identical functional properties prevailed*. The unsolved actuation riddle [...] creates the opposite problem—of explaining why language fails to change. (Weinreich et al., 1968, p.112, emphasis added)

However, as shown in Figure 4.2, rather than runaway harmonisation in all cases, what happens is that some runs appear to approach (not always monotonically) two stable plateaux, with lexicons having harmonic proportions in the neighbourhoods of 50% and 75%. The approach to these intermediate levels takes more or less time as a function of the magnitude of coarticulation.

Examination of the lexical contents of the agents across generations reveals that the [-GRAVE,-COMPACT]<sup>6</sup> vowel persists longer in the face of coarticulatory pressure toward low  $F_2$ , that is, the final-stage lexicons often have entries with 2 or 3 [+GRAVE] (low- $F_2$ ) vowels, and one or two [-GRAVE] (high  $F_2$ ) vowels. Essentially these are behaving like lexical transparent vowels. This can be explained by so-called “coarticulatory resistance”: experimental work has shown that some vowels, particularly high front vowels, are less prone to coarticulatory effects than others (Recasens, 1984; Beddor et al., 2002).<sup>7</sup> In the context of the current model, this is a straightforward consequence of the asymmetry of the distribution of the vowel prototypes.

Figure 4.3 gives an example of the distribution of vowels in the formant space at the beginning of one particular run of the simulation, although the layout is representative. The distribution of the vowels (after swapping the orientation of the axes) looks like an exaggerated version of a standard “vowel quadrilateral” as might be presented in an introductory phonology textbook. The asymmetry is clear, and the layout makes it obvious that the “coarticulatory resistance” of the [-GRAVE,-COMPACT] (low- $F_1$  high- $F_2$ ) vowel is a direct result of its being outside the average “reach” of coarticulation.

---

<sup>6</sup>[+HI,-BK] in traditional terms.

<sup>7</sup>Ohala (1994a) explicitly attributes the prevalence of high front neutral vowels in harmony to this resistance to coarticulation. It is not clear that the model as it is implemented here could result in opaque vowels, however.



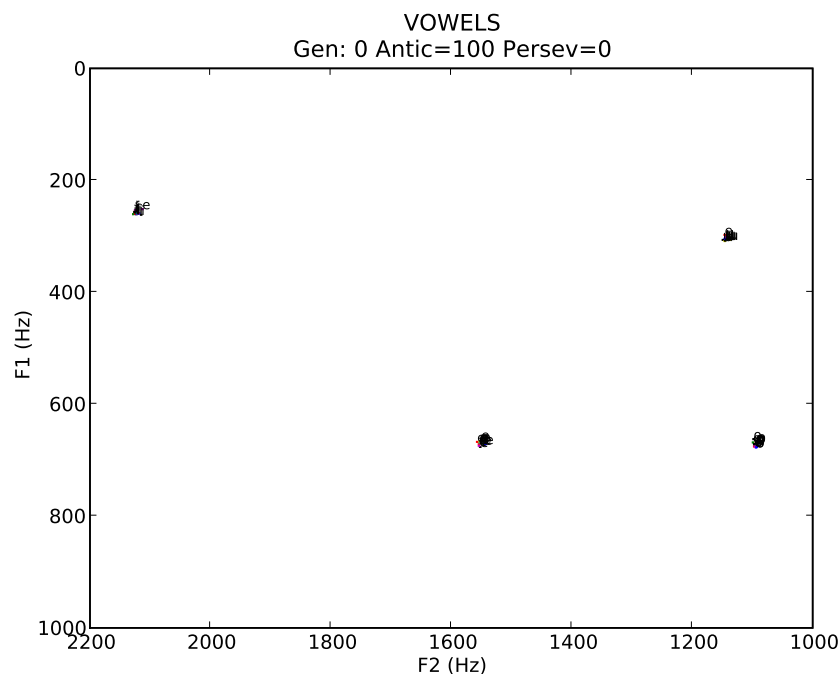


Figure 4.3: Sample distribution of vowels in  $F1 - F2$  parametric space.

In fact, Figure 4.3 makes it unclear how *any* harmony could arise, given (i) the  $F2$  separation between the high  $F1$  vowels (*approx.* 450Hz), and (ii) the amounts of coarticulation investigated here (*viz.* 400Hz max). In fact, without the addition of isotropic (Gaussian) noise to the output of the formant synthesis equations, there is no increase in harmony across time.<sup>8</sup> Thus, in the model presented here, at least, anisotropic (*i.e.* directionally-biased) noise—in the form of coarticulation—is not enough *by itself* to result

<sup>8</sup>There is typically a small jump in proportion of harmony after a single generation due to the fact that some proportion of the agents can be fairly extreme hypoarticulators.

in the diachronic emergence of vowel harmony, but it must be accompanied by some additional source of noise.

The synergistic interaction of the isotropic and anisotropic noise also accounts for the sporadic “jumps” in harmony, wherein a subset of the runs may “escape” the principal region of stability and end with a higher proportion of harmonic forms in the lexicon, or else reach a plateau much more quickly than other runs with the same parametric specification). This is clearly seen for five of the runs in Figure 4.2a. Depending on where a particular agent lies along the hypo/hyperarticulation continuum,<sup>9</sup> several “harmonisations”, which is to say misclassifications in the direction of increased harmony, may occur together within a generation and conspire to drive a language toward harmony much more quickly than is typical.

To the extent that the model captures the human situation, the distribution of harmonic languages across the world, namely their comparative rarity, would seem to suggest that whatever mechanism underlies dispersion in human vowels systems (*e.g.* synchronic or diachronic pressures toward contrast preservation or homophony avoidance), it is sufficiently strong to ensure that its effects typically outweigh the amounts of coarticulation characteristic of human speech.

---

<sup>9</sup>Recall that this is controlled by the shape parameters of the Beta distribution, which are normally distributed  $\alpha \sim \mathcal{N}(\mu = 40, \sigma = 5)$ ,  $\beta \sim \mathcal{N}(\mu = 5, \sigma = 1)$ .

**NON-MONOTONICITY**

The periodic, albeit rather dramatic, decreases in proportion of harmony that appear to occur roughly once per run above a certain threshold of coarticulation (approximately 125Hz) occurs due to the “diagonal proximity” of the [-GRAVE,+COMPACT] and [+GRAVE,-COMPACT] vowels, which is slightly more than half the distance between the [-GRAVE,-COMPACT] and [+GRAVE,-COMPACT] vowels. This is again a result of the isotropic noise added to the synthesised outputs: formant values are clipped “outside” the bounds of the quadrilateral, but allowed to vary freely within, and it is in the nature of stochastic noise that there will be outliers. To the best of my knowledge, this type of “cross-categorical” harmony variation does *not* occur in human language, although “height” harmony does exist and is presumably the result of systematic *F1* variation. It seems, then, that whatever the net result is of the sources of noise that I am collapsing together under a Gaussian here, it is constrained in ways that my model is not. This is a clear avenue for future research.

**4.2 CONCLUSIONS**

In this chapter, I focused on a particular instance of diachronic explanation: Ohala’s 1994b claim that vowel harmony emerges from the phonologisation of vowel-to-vowel coarticulation. Using a simple model of the

language transmission/acquisition feedback loop iterated over multiple generations, I showed how a gradient pattern of front/back coarticulation coupled with anisotropic noise from other factors (fatigue, noise, *etc.*) could eventually become phonologised as a categorical pattern of lexical harmony.

There remains much work to be done in fleshing out this model to more accurately reflect the conditions that obtain in real-world examples of sound change and phonologisation. The model as presented here fits broadly into the view of phonological diachrony espoused by Hale (2007), whereby sound changes and phonologisation are initiated within the heads of individual speaker-hearers. Of course, individuals acquire their language from multiple sources (hence more variable input forms), and children's language is often shaped as much by their peers as their parents (Labov, 1994), so even for a diachronician who ascribes to Hale's viewpoint, it seems unwise to avoid the influence of external actors. The incarnation of the model that is currently in development incorporates social networks, as well as acquisition from multi-source data (*cf.* p.83).

This chapter represents a first step in demonstrating that computational modelling can support—and even be a crucial component of—diachronic explanation of synchronic phonological patterns. Given the increasing push toward this style of explanation (*cf.* Blevins, 2004; Hale and Reiss, 2008), and the obstacles to empirically investigating phenomena which

arise over timescales potentially spanning centuries or millennia, the usefulness of computational modelling in putting these explanations on sound theoretical and empirical footing is clear.

#### 4.2.1 COARTICULATION, *prise deux*

The work presented in this chapter serves as an “existence proof” for the Ohalian theory of the origins of vowel harmony, namely that it results from the phonologisation of vowel-to-vowel coarticulation. However, there is a recent convergence of evidence that at least casts doubt on this view of the origins of vowel harmony, in particular with respect to the role played by coarticulation.

Beddor and colleagues (Beddor et al., 2002, 2007; Beddor, 2009) have recently demonstrated that both coarticulation—V-V and in VN sequences—and “perceptual compensation” for coarticulation are highly language-specific, in particular that anticipatory and perseverative coarticulation vary widely in degree across languages, and that compensation for coarticulation is largely attuned to a language’s amount of coarticulation. This immediately puts the “phonologisation of coarticulation” account, at least as it has been implemented here, on less certain footing. If listeners generally compensate as much as speakers tend to coarticulate, it is unclear whether failures of compensation happen frequently enough for phonologisation to gain any traction.

Independently of this, there is a line of research giving increasing evidence that language users have access to highly detailed episodes of linguistic experiences (Goldinger, 1996; Johnson, 1997b; Pierrehumbert, 2001; Hawkins, 2003, *inter alia*), and in particular that language users store acoustically-detailed, *non-segmented* exemplars of linguistic experiences (Silverman, 2006; Johnson, 2007; Välimaa-Blum, 2009). But if lexical representations are unsegmented acoustic trajectories, then there is no meaningful sense in which coarticulation, within words at least, happens at all. Consider the difference between the relatively palatalised [k] in *keep* versus the velarised [k] in *coop*, which are often highlighted as examples of anticipatory coarticulation. On the view presented in this thesis, this difference in place of articulation is not attributable to online, synchronic coarticulation, but instead is a product of the fact that these forms have only ever been heard by the language learner in their respective palatalised and velarised forms.

# CHAPTER 5

## DISCUSSION

### 5.1 PHONOLOGY IN COGNITIVE SCIENCE

Linguistics has always been considered one of the core disciplines of the cognitive sciences, from the earliest acknowledgement that a “cognitive revolution” was underway in psychology (Miller, 2003). Relatedly, Noam Chomsky has repeatedly made the claim that linguistics is (or should be construed as) a branch or subfield of cognitive psychology (Chomsky, 1980, 1986, 2000).<sup>1</sup> Notwithstanding this claim, much—perhaps most—of the research carried out in generative linguistics (arguably the contemporary mainstream, at least in North America) has been carried out in relative

---

<sup>1</sup>Chomsky often goes further in claiming that linguistics is “theoretical biology”, but he seems to mean this only in the sense that cognitive psychology, a high-level theory of how the brain organises humans’ interactions with the world, is also theoretical biology.

isolation from the other cognitive sciences, in particular ignoring one of the core debates in contemporary cognitive psychology: the nature and composition of mental/cognitive categories.

Modern linguists have historically favoured, and continue to rely on, Aristotelian mental categories defined in terms of necessary and sufficient conditions, typically expressed as categorical features/properties (see Halle and Bromberger, 1997, for an unusually explicit discussion and defense of this position). This stance appears to have been adopted by the generativists from the American structuralist tradition. The structuralists, being more or less behaviourist in disposition, made no claim—and in fact often argued against the possibility of making any claims—about the psychological status of the features they used, which simply signalled oppositions in abstract linguistic systems. The founders of the generative movement, explicitly interested in the psychological/mental aspects of human language, adopted the structuralists' features and essentially put them into the heads of human language users. This view of mental categories has all but disappeared in psychological theorising, where the debate over mental representations now centers on whether a prototype (Rosch, 1973, 1975) or exemplar-based (Medin and Schaffer, 1978) approach better accounts for the empirical facts.

One of the goals of this thesis has been to make the link between linguistics and cognitive psychology more explicit, by embracing the exemplar-based



approach to cognitive categories, one of the dominant approaches in modern cognitive psychology. The other dominant approach, prototype theory (Rosch, 1973, 1975), has been taken up in *Cognitive Grammar* (Langacker, 2008), an approach to human language which has historically focused less on structure, and more on how metaphor shapes the way humans use language. Empirically, it seems that the bulk of the evidence weighs in favour of exemplar-based approaches to categorisation, but there are still many unresolved issues and the debate is lively (Murphy, 2002). One set of data which argue clearly for an exemplar-based approach is so-called “gang effects”, in which a cluster of outliers can conspire to override a category prototype (*cf.* Chandler, 2002, for clear discussion). It seems to me that many of the recent “artificial grammar” learning experiments could be insightfully analysed as instances of gang effects overriding native linguistic competencies, but this remains to be seen.

Although there is already a significant literature in exemplar-based approaches to linguistics, enough that *The Linguistic Review* devoted a recent special issue to them (Gahl and Yu, 2006), few of the people working on these models have made it explicit that this approach brings linguistics more fully into the fold of the cognitive sciences. In addition to the ties to psychology mentioned in the preceding paragraphs, there are clear links to artificial intelligence/machine learning as well, as seen in subsection 2.3.3. In the next section, I claim that these models potentially open the door to more collaboration with neuroscience, as well, although—as is often the

case in attempts to integrate neural and cognitive approaches—it would be premature to claim anything more than an open door.

## 5.2 PHONOLOGY AS COGNITIVE SCIENCE

Having motivated the approach developed in this thesis with respect to a rapprochement between phonology (linguistics, more generally) and adjacent disciplines in the cognitive sciences, I will now discuss what it means for phonology to *be* a cognitive science, and how the approach developed here contributes to this.

### 5.2.1 THE MARRIAN APPROACH

In the opening chapter of his book *Vision* (Marr, 1982, p.19–27), David Marr laid out what would become one of the core organising principles of modern cognitive science: the investigation and explanation of information processing systems at multiple levels of analysis. On this view, a complete understanding (*viz.* description and explanation) of any information-processing system—by this point, it was standardly accepted that cognition was best described/explained in terms of information-processing (Neisser, 1967)—can only be achieved by investigation at multiple levels of analysis, or abstraction. In particular, Marr posited that (at least) a tri-level approach was necessary:

**The computational level** What is the function being computed by the system?

**The algorithmic level** Which algorithms (*viz.* representations and operations) are being used to compute this function?

**The implementational level** How are the representations and operations of the algorithm implemented in a physical substrate?

Marr was at pains to emphasise the fluidity of the relations between levels, noting:<sup>2</sup>

Each of the three levels of description will have its place in the eventual understanding of [ . . . ] information processing, and of course they are logically and causally related. But an important point to note is that since the three levels are only rather loosely coupled, some phenomena may be explained at only one or two of them. This means, for example, that a correct explanation of some psychophysical observation must be formulated at the appropriate level. (Marr, 1982, p.25)

In the context of the work presented in this thesis, and in particular the preceding discussion in this chapter, the question is whether phonology can or should be analysed from the Marrian stance. Though the issues are not necessarily clear-cut, and the project—assuming it is a valid one—is

---

<sup>2</sup>Cf. the discussion by Hardcastle (1996) highlighting the importance of “bridge laws” in cognitive science, for a contrasting view.

vast, I address some of these questions in the following sections.

### THE COMPUTATIONAL LEVEL

Whether expressed derivationally or in terms of constraints, whether with one, two, or several “strata” of computation, generative phonology has typically been seen as performing some sort of input-output mapping, or more generally as relating two sets of entities. In this sense, traditional phonology is clearly a computational-level account of the phonological component of the human faculty of language, viewed as an information-processing system. Moreover, it is by now well-known that both Optimality Theory (Prince and Smolensky, 2004) and derivational, or “rule-based” phonology (derivatives of Chomsky and Halle, 1968, broadly construed), the two major approaches to modern generative phonology, are—with only slight modifications—reducible to a common computational framework, finite-state transducers or regular relations (Kaplan and Kay, 1994; Bird and Ellison, 1994; Karttunen, 1998).

Notwithstanding this, and despite clearly articulated warnings about this kind of mistake (Kaplan, 1987), many generative phonologists have historically viewed, and continue to view, their theories as being situated at the algorithmic level of the Marrian hierarchy. For proponents of derivational accounts, it may be that the language of the framework, where analyses are described in terms of the procedural application of ordered rules, is

what leads researchers astray (*cf.* the discussion in Isac and Reiss, 2008, ch. 4). In other cases, it may simply be a result of a confusion of levels, as in Purnell et al. (2010), where an account in terms of finite-state automata is given, and then explicitly claimed to exemplify an algorithmic level analysis. Finite-state automata are computational/mathematical abstracta, distinct from their algorithmic implementations.

### THE ALGORITHMIC LEVEL

The model of phonology presented in this thesis is, I claim, an algorithm level account of the acquisition and productive, generalised use of vowel harmony (and by hypothesised extension other “phonetically natural” assimilatory phenomena). The specification of LIBPHON as a theory/model of phonology requires explicit attention to representations and well-defined operations over those representations.

More generally, I speculate that the class of exemplar/instance-based approaches to *phonological*<sup>3</sup> problems, at least those that are specified enough to be implemented as models, is the only viable family of algorithm level accounts on the table.

---

<sup>3</sup>As distinct from *e.g.* phonetics or speech perception, assuming such a distinction stands.

**THE IMPLEMENTATION LEVEL**

I will have little to say about the “lowest” level of the Marrian hierarchy, other than to point out that in the realm of computational neuroscience (distinct from and not to be confused with *connectionism*, the variety of “neural” modelling more typically familiar to cognitive scientists), there are existing biologically-plausible models that seem to implement something very close to exemplar-type memory systems, *e.g.* the family of *Adaptive Resonance Theory* (ART) models developed by Stephen Grossberg and his students (Grossberg, 2003), and Frank Guenther’s *DIVA* model (Guenther, 2001).

Of course, the model presented here is incomplete in many ways, as is our current understanding of the neural underpinnings of speech perception and production and higher-level cognitive representations of language, so it is premature to do anything more than speculate at this point. Moreover, as highlighted above, the levels of the Marrian hierarchy are only loosely coupled, and thus it may be the case that phonology, as distinct from phonetics and speech perception, is only to be accounted for at the computational and algorithmic levels.

### 5.2.2 THE CONTINUED NECESSITY OF “TRADITIONAL” PHONOLOGY

There is a common perception among phonologists that instance-based accounts are posed as *replacements* of traditional phonological analyses, and indeed some exemplar theorists clearly espouse this position (Silverman, 2006; Kirchner and Moore, 2009). Taking seriously the Marrian multi-level view leads us to conclude that this position is likely a mistaken one, and indeed that is what I claim here.

A purely instrumentalist motivation for continuing to carry out computational level phonological analyses is that, for the moment at least, it remains exceedingly difficult to “do phonology” in the framework developed here, especially the hypothetical future version of this framework that fully-incorporates the temporal variability of real speech. One of the motivations for exemplar-based approaches to phonology is the set of recent and ongoing discoveries concerning the role that so-called “fine phonetic detail” plays in controlling systematically produced and perceived aspects of speech production (Hawkins, 2003). It seems sometimes that nearly every property of speech that systematically varies is or can be used by listeners to recover or interpret parts of the perceived signal. But this leads to (i) the problem of a surfeit of data with, as yet, no clear model of how to usefully search it, and (ii) an implicit rejection of at least part of the basic methodology of science, which assumes that some data are better

than others in explaining phenomena of interest. The “coarse-grainings” and idealisations made in computational level phonology make it possible for us as researchers to get a handle on our data, which in turns gives us a usable wedge into a potential instance-based account. Of course, the finer-grained analyses that instance-based approaches yield will sometimes reveal that particular computational level idealisations can or should be profitably relaxed; the coupling between levels is bidirectional.

The whole point of Marr’s account is that all levels of analysis are necessary for a complete understanding of some phenomenon, and he made it clear that computational level accounts, in particular, are profitable to pursue, even praising the successes of generative syntax in this regard:

Chomsky’s 1965 theory of transformational grammar is a true computational theory in the sense used earlier. It is concerned solely with specifying what the syntactic decomposition of an English sentence should be, and not at all with how that decomposition should be achieved. Chomsky himself was very clear about this—it is roughly his distinction between competence and performance, though his idea of performance did include other factors, like stopping in midutterance—*but the fact that his theory is defined by transformations, which look like computations, seems to have confused many people.* (Marr, 1982, p.28; emphasis added)



### 5.2.3 CATEGORICITY AND GRADIENCE IN PHONOLOGY

Claims about categorical versus gradient distinctions are a hot topic in phonology right now, especially with the rise of eliminativist exemplarism, but they abound in the *Laboratory Phonology* community as well. Traditionally, categorical phenomena were taken to be a hallmark of phonology, while gradience was indicative of phonetic processes.

This thesis makes two implicit claims about sources of categorical distinctions in phonology. On the first hand, some categoricity results from the coarse graining that computational level accounts impose on the data in order to make initial sense of it and to find high-level generalizations. This categoricity is contingent and model-dependent, and open to revision if some idealisation should prove to be profitably reversible. A second, and potentially more interesting, source of categoricity stems from the categorical (and compositional) nature of semantics/meaning. The implementation of LIBPHON crucially assumes that gradient sound differences do NOT result in gradient “shades of meaning”: an instance is categorised or labelled one way or another, or maybe even cross-categorised, but crucially cannot be “in-between”. I take this to be fairly uncontroversial although it is presumably an empirical question.

### 5.3 DIACHRONIC PHONOLOGY AS PREDICTIVE SCIENCE

One of the advantages of the computational evolutionary approach pursued in chapter 4 is that it becomes possible not only to make retrodictions about the distribution of phonological patterns, but also to make predictions. By starting from a known state and running a simulation, a “possible future” is generated. Iterated many times, this generates a distribution over “possible futures” that in principle licenses statements of the form *under the set of circumstances laid out in this set of simulations, vowel harmony will develop from an initially non-harmonic state X% of the time.*<sup>4</sup>

Of course, this is a gross simplification, given the acknowledged role in language change that social factors not even broached in this thesis play, for example, language contact scenarios (with or without accompanying socio-political imbalances), diffusion of particular speech styles of important people via media, *etc.* Many of these factors are studied by sociolinguists, and are in principle amenable to sufficient analysis that an estimate of their parametric values can be incorporated into the type of diachronic model developed here. With a sufficient degree of caution, it seems likely that statistical *short-term* predictions can be made, *e.g.* statements of the form *under*

---

<sup>4</sup>According to current philosophy of science, these kinds of statistical generalisations under certain circumstances have the same status as traditional “laws of nature” (Salmon, 1998).

*the set of circumstances laid out in this set of simulations, over  $X$  generations, the distribution of pattern  $Y$  will move  $Z$  distance along dimension  $W$  with a  $V\%$  confidence interval.* This highlights the massively interdisciplinary nature of phonology as a potentially explanatory<sup>5</sup> science, requiring input from phoneticians, phonologists, sociolinguists and computational modellers, among many others.

## 5.4 LEARNING

In the spirit of “ideal observer” analysis (Geisler, 2003), the agents in chapter 4 are interpretable as Bayesian learners whose goal is to infer an appropriate mental model on the basis of the data to which they are exposed. This inference task is encapsulated in Bayes’s Theorem:

$$P(H = h|D = d) = \frac{P(D = d|H = h)P(H = h)}{P(D = d)}$$

where  $h$  is a candidate hypothesis from the space of possible hypotheses,  $\mathcal{H}$ , and  $d$  is the data that are available to an agent. In words, Bayes’s Theorem assigns a posterior probability to each candidate hypothesis proportional to the likelihood that the data in question would be generated by that hypothesis and the prior probability of that hypothesis. Note that Bayes’s Theorem by itself is not enough to perform inference, as it only assigns

---

<sup>5</sup>Explanation being the backward-looking cousin of prediction (Salmon, 1998).

the probabilities to the hypotheses, and gives no indication about how (or whether) to select a given hypothesis. As stated earlier, I allow my agents to entertain a single grammar at any given moment,<sup>6</sup> and so my agents will perform *maximum a posteriori* (MAP) inference, selecting the hypothesis with the highest posterior probability as the correct one.

Although humans are arguably not optimal learners in the Bayesian sense, ideal observer analysis uses the concepts of Bayesian statistical inference to determine optimal performance on a particular task, given the existing constraints on the agent.

The purpose of deriving an ideal observer is to determine the optimal performance in a task, given the physical properties of the environment and stimuli. Organisms generally do not perform optimally, and hence one should not think of an ideal observer as a potentially realistic model of the actual performance of the organism. Rather, the value of an ideal observer is to provide a precise measure of the stimulus information available to perform the task, a computational theory of how to perform the task, and an appropriate benchmark against which to evaluate the performance of the organism. (Geisler, 2003)<sup>7</sup>

Current research in psychology points to increasing evidence for Bayesian

---

<sup>6</sup>See *e.g.* Yang (2002) for an example of the alternative view.

<sup>7</sup>Note the relation here to one aspect of Epstein's (2006) discussion of *bounded rationality* in agent-based models, which may be extrinsic, as a bound on the information available to agents, or intrinsic, as a bound on agents' computing power (or, of course, both)

reasoning in human cognition. The articles in (Chater and Oaksford, 2008) document several areas of cognitive science in which human feats of cognition at both high and low levels (*e.g.* vision and decision-making) are best modelled by Bayesian methods. In the context of modelling phonological acquisition in particular, much research over the past two decades (Ellison, 1992; Hayes and Londe, 2006; Goldwater and Johnson, 2004) is converging on probabilistic (Bayesian and information-theoretic) models as being a good account of human language learning.<sup>8</sup>

With respect to human acquisition of vowel harmony, the data that are available (Leiwo et al. (2006) for Finnish, Altan (2007) for Turkish, MacWhinney (1978) for Hungarian) suggest that children very rarely make harmony mistakes, and they have full early mastery of the phenomenon. While this in and of itself is no argument for language learners as “ideal observers”, the facility and accuracy with which children acquire vowel harmony suggests that the assumption of near-optimal learning is at least plausible.

---

<sup>8</sup>This is not specifically a claim that is meant to bear on the issue of innateness, although it is often proposed in that context. At the very least, it appears that a probabilistic approach best accounts for the learning of those aspects of language which are indubitably learned (*e.g.* any language-specific properties).

## CHAPTER 6

### CONCLUSIONS AND FUTURE

### DIRECTIONS

This thesis has been a computational investigation of the population-level and speaker-level emergence of vowel harmony.

I began with an instance-based model of the synchronic acquisition of vowel harmony, with the goal of investigating whether an acoustic word-based exemplar approach could capture the relevant phenomenological aspects of stem-controlled vowel harmony, including opaque and neutral transparency. On the basis of a limited quantity of input data, the model, `LIBPHON`, learned to produce unseen forms with harmonically correct affixes, including patterns of vowel neutrality. This was followed with a model whose goal was to identify conditions under which lexical vowel

harmony could arise in a population of speakers from an initially non-harmonic state, essentially seeking some form of evidence for or against John Ohala's 1994b theory that vowel harmony arises through the phonologisation of vowel-to-vowel coarticulation. This iterated learning model showed that harmony could arise from local, stochastically-manifested vowel-to-vowel coarticulation, thereby providing a plausibility argument for the Ohalian account. Moreover, it showed that even in the absence of external "social" factors (*e.g.* language contact, the influence of prestige dialects, *etc.*), the march toward harmony is not inevitable in the face of steady coarticulatory pressures. In particular, some "intermediate" degrees of harmony, in which a strict subset of the lexicon has harmonised, were stable over many (up to hundreds) generations of transmission and acquisition. As discussed in chapter 4, this goes some way toward addressing the "actuation problem" of Weinreich et al. (1968), in demonstrating that "why a change happens when it does" is the result of interaction between the anisotropic pressure of coarticulation and the presumed isotropic influence of the sum of all other noise sources (internal and external).

## 6.1 CONTRIBUTIONS

The core contributions of the thesis are the simulations in chapter 3 and chapter 4, which together constitute "sufficiency" proofs for (i) the viability

of historical/diachronic explanations of synchronic phonological patterns, in particular the Ohalian theory of the origins of vowel harmony (*viz.* that it emerges from the phonologisation of vowel-to-vowel coarticulation), and (ii) the viability of lazy learning (*qua* nonparametric regression) over numerical representations as a framework for phonological acquisition, in particular of a general, productive pattern of synchronic alternations, arguing directly against claims that such patterns necessitate the positing of structured, symbolic representations.

The simulations in chapter 3 constitute the beginnings of a novel account of vowel harmony, illustrating how insights from machine learning can relatively straightforwardly derive productive, general patterns, and obviating the *ad hoc* theoretical machinery that has been required in order to deal with neutral vowels. Moreover, LIBPHON is a relatively novel type of generative instance-based system, drawing on the advantages of both kernel-based and nearest-neighbour approaches to regression; this combination of approaches is understudied in the statistics and machine learning literature and opens up interesting avenues for further research, *e.g.* determining the relative importance of distance/similarity measure versus neighbourhood size as well as the overall shape of the kernel function.

Empirically, the simulations in chapter 4 have the potential to make falsifiable predictions about the typological distribution of assimilatory patterns. As our scientific understanding of articulatory-perceptual, cognitive, and



social biases increases, it should in principle be possible to incorporate this information into iterated learning models, both with respect to the distribution over inter-agent interactions and the internal state and comprehension/production modules, and make concrete predictions about the time-course and shape of phonological change in specific cases, at least on short timescales (see Janda and Joseph, 2003).

On a more conceptual level, the thesis makes some progress on developing stronger connections between more narrowly linguistic phonology and cognitive science, adopting an explicitly Marrian stance toward phonological explanation, and drawing on theories and methods from adjacent disciplines (in particular psychology and AI/machine learning).

## 6.2 REMAINING WORK

### 6.2.1 ACQUISITION AND USE

#### VOWEL HARMONY

As attested by the sheer volume of literature devoted to the subject, the account of vowel harmony developed here only scratches the surface of the phenomena it comprises. For a complete account of the acquisition and use of vowel harmony, LIBPHON will need to be tested on dominant-recessive data, consonantal interactions (*e.g.* palatalised consonants in

Turkish, which interact in non-trivial ways with palatal harmony), and “non-systematic” exceptions, *i.e.* the effect on acquisition of disharmonic stems in a lexicon.

## PHONOLOGY

Beyond the domain of vowel harmony, there are Goldsmith’s 2009 “kernels” (see section 1.1), those atheoretic phonological generalisations that define the body of evidence that phonologists (tacitly) agree form the core problems of the discipline, problems which any theory aspiring to completeness must account for. Goldsmith mentions generalisations about sound inventories (*e.g.* considerations of symmetry, typological facts) and stress systems,<sup>1</sup> among others.

To this we may add the “Bach test” originally proposed by Lise Menn (Halle, 1978), wherein native English speakers pluralise (or form the possessive) of *Bach* [bax] as [baxs]. This is meant to be illustrative of how competent language users generalise in the face of data outside their experience.

---

<sup>1</sup>Jackendoff (2002) singles out stress systems as good targets for computational modelling, and there has in fact been some work in this area, *e.g.* Dresher and Kaye (1990); Daelemans et al. (1994).

### 6.2.2 CHANGE

An immediate avenue for further exploration is to apply the model in chapter 4 to the mixed modes of transmission discussed in chapter 3 (*i.e.* oblique flow and peer interactions) to see whether these affect the rate of phonologisation, or ultimate levels of harmony (*e.g.* Do the intermediate levels still occur? Are they shifted?). Hudson Kam and Newport (2005) give evidence that children tend to regularise variable inputs more than adults,<sup>2</sup> which perhaps leads to the prediction that peer interactions will speed phonologisation by smoothing isotropic variation, but tending to leave anisotropic variation intact. More generally, it is by now accepted that usage-related factors beyond the usual parameters of acquisition, such as word frequencies and social graph connectivity (especially with respect to prestigious speakers) play a large role in language change (Weinreich et al., 1968; Labov, 1994; Bybee, 2001). Incorporating frequency data and exploring the effects of varying communicative links in a social network are logical next steps in assessing the model of change presented here.

Related to this is the set of phenomena gathered under the rubric *creolisation*: previous modelling work (Steels, 1997; Oudeyer, 2006, *inter alia*) has shown that horizontal information flow results in the self-organization of a shared lexicon/communicative code. Previous empirical work (Hudson Kam and Newport, 2005) has shown that children tend to regularise

---

<sup>2</sup>This makes sense on the view of acquisition provided by LIBPHON; sparsely-populated instance spaces result in high bias and low variance, which “smooths” variable inputs.

variability in their input. In combination, these facts point to the model introduced here being a plausible account of the pidginisation/creolisation cycle, in particular with horizontal information flow at the *adult* level in a language contact scenario, in conjunction with vertical or oblique flow with peer interactions (*cf.* DeGraff, 2009, for discussion of this scenario).

### 6.2.3 LIBPHON

In section 3.1 I have already discussed some of LIBPHON’s shortcomings. The key avenue of development here will be to determine the best way of working with real speech, in particular with temporally-variable instances of particular LABELS. It strikes me that dynamic time warping, notwithstanding its cognitive implausibility and relative inefficiency, could be used as a first pass, if only to assess questions like *can LibPhon’s general architecture work with this kind of data?* As discussed, the algorithm is computationally inefficient (quadratic in time and space), which will lead to problems in the face of richer (*viz.* higher dimensionality) acoustic representations, *e.g.* encoding based on mel frequency cepstral coefficients or linear predictive coding (Johnson, 2003), or even full spectrograms. I suggested that a modified version of Stevens’s 2002 “acoustic landmark” based account of lexical access might be a way to escape dealing with temporal variability, and this seems like another viable avenue for further research.

On the theoretical side, a more thorough investigation of the properties of LIBPHON's generation algorithm are in order. Although there is a vast literature on distance/similarity-based approaches to learning (commonly called *kernel*-based methods), and nearest-neighbour learning, there has been very little work on algorithms which combine these approaches, and how such blended methods fare in comparison with their individual parts remains to be assessed.

# Appendices

# APPENDIX A

## EQUATIONS FOR FORMANT

### SYNTHESIS

The following equations are from de Boer (2001). They are used to synthesize formant values from real-valued specifications of [COMPACT], [GRAVE] and [FLAT] (roughly, *SPE ROUND*) articulatory feature values. Since I did not use the latter in my model, it was consistently set to zero. Also, I present only the equations for  $F1$  and  $F2$  here; de Boer (2001) used  $F1 - F4$  in his simulations.

$$\begin{aligned}
F1 = & ((-392 + 392f)c^2 + (596 - 668f)c + (-146 + 166f))g^2 \\
& + ((348 - 348f)c^2 + (-494 + 606f)c + (141 - 175f))g \\
& + ((340 - 72f)c^2 + (-796 + 108f)c + (708 - 38f))
\end{aligned}$$

$$\begin{aligned}
F2 = & ((-1200 + 1208f)c^2 + (1320 - 1328f)c + (118 - 158f))g^2 \\
& + ((1864 - 1488f)c^2 + (-2644 + 1510f)c + (-561 + 221f))g \\
& + ((-670 + 490f)c^2 + (1355 - 697f)c + (1517 - 117f))
\end{aligned}$$



## **APPENDIX B**

### **BETA DISTRIBUTION**

The Beta distribution models events constrained to happen within a bounded interval. It has two shape parameters,  $\alpha, \beta$ , whose general effect are shown in B.1. It is used here to model the hypo/hyperarticulation continuum (Lindblom, 1990).

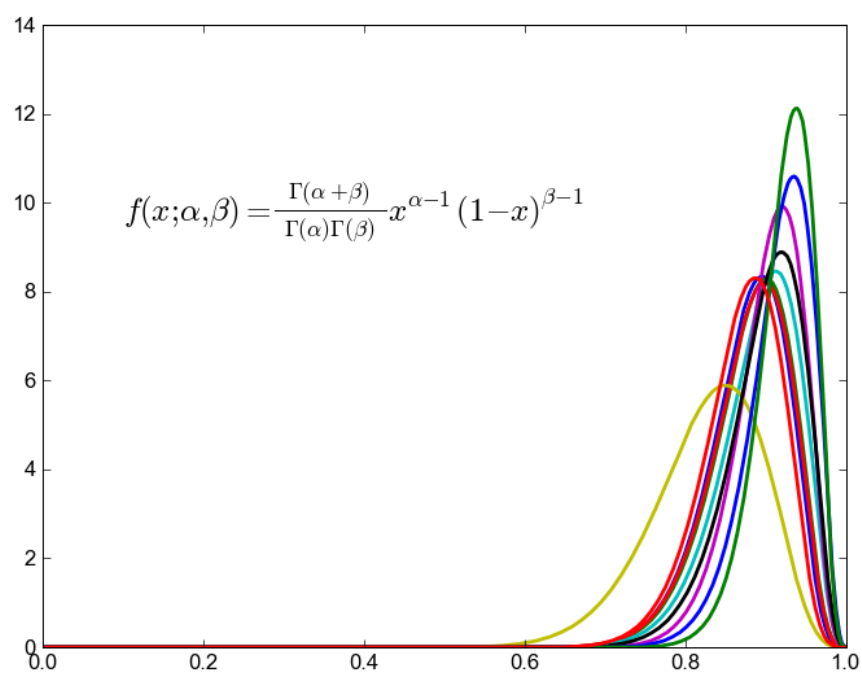


Figure B.1: Beta distribution for ten values of  $\alpha \sim \mathcal{N}(\mu = 40, \sigma = 5), \beta \sim \mathcal{N}(\mu = 5, \sigma = 1)$ .

# APPENDIX C

## *k*-MEANS CLUSTERING

The *k*-means algorithm is an iterative algorithm for partitioning  $n$  data points in an  $m$ -dimensional space into  $k$  clusters,  $C = \{C_1, \dots, C_k\}$ . This is achieved by minimizing *total intra-cluster variance* (MacKay, 2002):

$$V = \sum_{i=1}^k \sum_{x_j \in C_i} (x_j - \mu_i)^2$$

The standard algorithm begins with  $k$  points in the instance-space (which may be initialised randomly, or as data points), then iteratively assigns each input datum to the closest such point, and recalculates the new means for each cluster as the centroid of the points assigned to that cluster.

## BIBLIOGRAPHY

- Aha, D. (1997). Lazy learning. In *Lazy Learning*, pages 7–10. Kluwer Academic Publishers.
- Aha, D., Kibler, D., and Albert, M. (1991). Instance-based learning algorithms. *Machine Learning*, 6:37–66.
- Albright, A. and Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90:119–161.
- Altan, A. (2007). The acquisition of vowel harmony in Turkish. Poster presented at OCP4.
- Aoki, H. (1966). Nez Percé vowel harmony and proto-Sahaptian vowels. *Language*, 42:759–767.
- Archangeli, D. and Pulleyblank, D. (1994). *Grounded Phonology*. MIT Press.
- Ashby, F. G. and Alfonso-Reese, L. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, 39(2):216–233.

- Bakovic, E. (2000). *Harmony, Dominance and Control*. PhD thesis, Rutgers University.
- Bakovic, E. (2003). Vowel harmony and stem identity. Technical Report ROA-540, UC San Diego.
- Baudouin de Courtenay, J. N. (1895). *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics*. Indiana University Studies in the History and Theory of Linguistics. Indiana University Press. Edited and translated by E. Stankiewicz (1972).
- Beckman, M. and Edwards, J. (2000). The Ontogeny of Phonological Categories and the Primacy of Lexical Learning in Linguistic Development. *Child Development*, 71:240–249.
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85:785–821.
- Beddor, P. S., Brasher, A., and Narayan, C. (2007). Applying perceptual methods to the study of phonetic variation and sound change. In Solé, M.-J., Beddor, P. S., and Ohala, M., editors, *Experimental Approaches to Phonology*, pages 125–143. Oxford University Press.
- Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30:591–627.

- Beesley, K. and Karttunen, L. (2003). *Finite-State Morphology*. CSLI Studies in Computational Linguistics. CSLI Publications.
- Berko Gleason, J. (1958). The Child's Learning of English Morphology. *Word*, 14:150–177.
- Bird, S. and Ellison, T. M. (1994). One-level phonology: autosegmental representations and rules as finite automata. *Computational Linguistics*, 20:55–90.
- Blevins, J. (2004). *Evolutionary Phonology*. Cambridge University Press.
- Bregman, A. (1990). *Auditory Scene Analysis*. MIT Press.
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge University Press.
- Carroll, J. (2004). Letter knowledge precipitates phoneme segmentation, but not phoneme invariance. *Journal of Research in Reading*, 27:212–225.
- Chandler, S. (2002). Skousen's analogical approach as an exemplar-based model of categorization. In Skousen, R., Lonsdale, D., and Parkinson, D., editors, *Analogical Modeling*, pages 51–108. John Benjamins.
- Chater, N. and Oaksford, M., editors (2008). *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. Oxford University Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, N. (1966). *Cartesian Linguistics: A chapter in the history of rationalist thought*. Harper and Row.

- Chomsky, N. (1980). *Rules and Representations*. Columbia University Press.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origins and Use*. Praeger.
- Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.
- Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper and Row.
- Choudhury, M. (2007). *Computational Models of Real World Phonological Change*. PhD thesis, Indian Institute of Technology, Kharagpur, India.
- Choudhury, M., Basu, A., and Sarkar, S. (2006). Multi-Agent Simulation of the Emergence of Schwa Deletion Pattern in Hindi. *Journal of Artificial Societies and Social Simulation*, 9. Retrieved from <http://jasss.soc.surrey.ac.uk/9/2/2.html> on 2010-07-16.
- Chouinard, M. and Clark, E. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30:637–669.
- Clements, G. N. (1980). *Vowel harmony in nonlinear generative phonology: an autosegmental model*. Indiana University Linguistics Club.
- Cole, J. (2009). Emergent feature structures: harmony systems in exemplar models of phonology. *Language Sciences*, 31. Indiana University Linguistics Colloquium, Sept. 2007.

- Cole, J. and Kisseberth, C. (1994). An optimal domains theory of harmony. *Studies in the Linguistic Sciences*, 24:1–13.
- Coleman, J. (1998). Cognitive reality and the phonological lexicon: A review. *Journal of Neurolinguistics*, 11:295–320.
- Coleman, J. (2000). Where is coarticulation? In Broe, M. and Pierrehumbert, J., editors, *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, pages 102–117. Cambridge University Press.
- Coleman, J. (2002). Phonetic representations in the mental lexicon. In Laks, B. and Durand, J., editors, *Phonetics, Phonology, and Cognition*, chapter 4. Oxford University Press.
- Cooper, F., Delattre, P., Liberman, A., Borst, J., and Gerstman, L. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 24:597–606.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27.
- Daelemans, W., Gillis, S., and Durieux, G. (1994). The acquisition of stress: A data-oriented approach. *Computational Linguistics*, 20:421–451.
- de Boer, B. (2001). *The Origins of Vowel Systems*, volume 1 of *Studies In The Evolution of Language*. Oxford University Press.
- DeGraff, M. (2009). Language acquisition in creolization and, thus, lan-



- guage change: Some cartesian-uniformitarian boundary conditions. *Language and Linguistics Compass*, 3:888–971.
- Démonet, J.-F., Thierry, G., and Nespoulos, J.-L. (2002). Towards imaging the neural correlates of language functions. In Laks, B. and Durand, J., editors, *Phonetics, Phonology, and Cognition*, pages 244–253. Oxford University Press.
- Deng, L. and O’Shaughnessy, D. (2003). *Speech Processing: a dynamic and optimization-oriented approach*, volume 17 of *Signal Processing and Communications*. CRC Press.
- Denis, D. (2010). Passive diagnostics of contrast. Presented at Montreal-Ottawa-Toronto Phonology Workshop.
- Dras, M. and Harrison, K. D. (2002). Emergent Behavior in Phonological Pattern Change. In Standish, R. K., Bedau, M. A., and Abass, H. A., editors, *Artificial Life VIII*, pages 390–393. Oxford University Press.
- Dresher, B. and Kaye, J. (1990). A computational learning model for metrical phonology. *Cognition*, 34:137–195.
- Ellison, T. M. (1992). Discovering vowel harmony. In Daelemans, W. and Powers, D., editors, *Background and Experiments in Machine Learning of Natural Language*. ITK.
- Epstein, J. M. (2006). *Generative Social Science: Studies in agent-based compu-*

- tational modeling*. Princeton Studies in Complexity. Princeton University Press.
- Ettlinger, M. (2007). An exemplar-based model of chain shifts. In *Proceedings of the 16th International Congress of the Phonetic Science*, pages 685–688.
- Faber, A. (1992). Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing. In Downing, P., Lima, S., and M. N., editors, *The Linguistics of Literacy*, pages 111–135. John Benjamins.
- Ferguson, C. and Farwell, C. (1975). Words and sounds in early language acquisition. *Language*, 51:419–439.
- Fix, E. and Hodges, J. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine.
- Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14:3–28.
- Gafos, A. (1999). *The articulatory basis of locality in phonology*. PhD thesis, New York University.
- Gahl, S. and Yu, A., editors (2006). *Exemplar-based models in linguistics (Special issue)*, volume 23. de Gruyter.
- Gasser, M. and Lee, C.-D. (1992). Networks that learn about phonological

- feature persistence. In *Connectionist natural language processing: readings from Connection Science*, pages 349–362. Intellect Books.
- Geisler, W. S. (2003). Ideal Observer analysis. In Chalupa, L. and Inner, J., editors, *The Visual Neurosciences*, chapter 52, pages 825–837. MIT press.
- Goldinger, S. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1166–1183.
- Goldsmith, J. (2009). Theory, kernels, data, methods. Presented at The 17th Manchester Phonology Meeting, Manchester, UK.
- Goldsmith, J. and Riggle, J. (2007). Information theoretic approaches to phonological structure: the case of Finnish vowel harmony. ms. University of Chicago.
- Goldsmith, J. and Xanthos, A. (2009). Learning phonological categories. *Language*, 85:4–38.
- Goldwater, S. and Johnson, M. (2004). Priors in Bayesian Learning of Phonological Rules. In *Proceedings of SIGPHON 2004*.
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31:423–445.
- Grünwald, P. (2007). *The Minimum Description Length Principle*. MIT Press.

- Guenther, F. (2001). Neural modeling of speech production. In *Proceedings of the 4th International Nijmegen Speech Motor Conference*.
- Hale, M. (2007). *Theory and Method in Historical Linguistics*. Oxford University Press.
- Hale, M. and Reiss, C. (2000). Phonology as Cognition. In Burton-Roberts, N., Carr, P., and Docherty, G., editors, *Phonological Knowledge*, pages 161–184. Oxford University Press.
- Hale, M. and Reiss, C. (2008). *The Phonological Enterprise*. Oxford University Press.
- Halle, M. (1978). Knowledge unlearned and untaught: what speakers know about the sounds of their language. In Halle, M., Bresnan, J., and Miller, G., editors, *Linguistic Theory and Psychological Reality*, pages 294–303. MIT Press.
- Halle, M. (1997). Some consequences of the representation of words in memory. *Lingua*, 100:91–100.
- Halle, M. and Bromberger, S. (1997). The content of the phonological sign: A comparison between their use in derivational theories and optimality theories. In Roca, I., editor, *Derivations and Constraints in Phonology*, pages 93–123. Clarendon Press.
- Halle, M., Vaux, B., and Wolfe, A. (2000). On Feature Spreading and the Representation of Place of Articulation. *Linguistic Inquiry*, 31:387–444.

- Hardcastle, V. (1996). *How to Build a Theory in Cognitive Science*. Suny Series in Philosophy and Biology. SUNY Press.
- Hare, M. (1990). The role of similarity in Hungarian vowel harmony: a connectionist account. *Connection Science*, 2:123–150.
- Hare, M. and Elman, J. (1995). Learning and morphological change. *Cognition*, 56:61–98.
- Harrison, K. D., Dras, M., and Kaun, A. (2003). Agent-based modeling of the evolution of vowel harmony. In *Proceedings of NELS32*.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *Elements of Statistical Learning*. Springer Series in Statistics. Springer-Verlag, 2 edition.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31:373–405.
- Hayes, B., Kirchner, R., and Steriade, D., editors (2004). *Phonetically Based Phonology*. Cambridge University Press.
- Hayes, B. and Londe, Z. C. (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology*, 23:59–104.
- Hertz, J., Krogh, A., and Palmer, R. (1991). *Introduction to the theory of neural computation*, volume 1 of *Santa Fe Institute Studies in the Sciences of Complexity*. Basic Books.
- Hickok, G. (2009). Eight problems for the mirror neuron theory of action

- understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21:1229–1243.
- Hintzman, D. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93:411–428.
- Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In *Current Progress in Historical Linguistics*. NorthHolland.
- Hruschka, D., Christiansen, M., Blythe, R., Croft, W., Heggarty, P., Mufwene, S., Pierrehumbert, J., and Poplack, S. (2009). Building social cognitive models of language change. *Trends in Cognitive Sciences*, 13:464–469.
- Hudson Kam, C. and Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1:151–195.
- Hyman, L. (1972). Nasals and nasalization in Kwa. *Studies in African Linguistics*, 4:167–206.
- Isac, D. and Reiss, C. (2008). *I-language: An introduction to linguistics as cognitive science*. Oxford Core Linguistics. Oxford University Press.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

- Jäkel, F., Schölkopf, B., and Wichman, F. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review*, 15(2):256–271.
- Janda, R. and Joseph, B. (2003). On language, change, and language change – or, of history, linguistics, and historical linguistics. In Joseph, B. and Janda, R., editors, *The Handbook of Historical Linguistics*, pages 3–180. Blackwell.
- Joanisse, M. (2000). *Connectionist Phonology*. PhD thesis, University of Southern California.
- Johnson, C. D. (1972). *Formal Aspects of Phonological Description*. Mouton.
- Johnson, K. (1997a). The auditory/perceptual basis for speech segmentation. *OSU Working Papers in Linguistics*, 50:101–113.
- Johnson, K. (1997b). Speech perception without speaker normalization: an exemplar model. In *Talker Variability in Speech Processing*, chapter 8, pages 145–166. Academic Press.
- Johnson, K. (2003). *Acoustic and Auditory Phonetics*. Wiley-Blackwell, 2 edition.
- Johnson, K. (2007). *Decision and Mechanisms in Exemplar-based Phonology*, chapter 3, pages 25–40. Oxford University Press.
- Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends*

*in Cognitive Sciences*, 3:323–328.

Jusczyk, P. (2000). *The Discovery of Spoken Language*. MIT Press.

Kaplan, R. (1987). Three seductions of computational psycholinguistics. In *Linguistic Theory and Computer Applications*. CSLI Publications.

Kaplan, R. and Kay, M. (1981). Phonological rules and finite-state transducers. [Abstract] Linguistic Society of America Meeting Handbook.

Kaplan, R. and Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, 20:331–378.

Karttunen, L. (1993). Finite-state Constraints. In Goldsmith, J., editor, *The Last Phonological Rule*, Studies in Contemporary Linguistics. University of Chicago Press.

Karttunen, L. (1998). The Proper Treatment of Optimality in Computational Phonology. In *Proceedings of the International Workshop on Finite State Methods in Natural Language Processing*, pages 1–27. Association for Computational Linguistics.

Keating, P. (1985). Universal phonetics and the organization of grammars. In Fromkin, V., editor, *Phonetic Linguistics: Essays in honor of Peter Ladefoged*. Academic Press.

Kertész, Z. (2003). Vowel harmony and the stratified lexicon of Hungarian. In *The Odd Yearbook* 7. ELTE Press.



- Kingston, J. (2007). The phonetics-phonology interface. In *The Cambridge Handbook of Phonology*. Cambridge University Press.
- Kiparsky, P. (1968). How abstract is phonology? In Kiparsky, P., editor, *Explanation in phonology*, pages 119–164. Foris.
- Kiparsky, P. (2008). Universals constrain change; change results in typological generalizations. In Good, J., editor, *Linguistic Universals and Language Change*, pages 23–53. Oxford University Press.
- Kirby, S. (1996). *Function, Selection and Innateness: the Emergence of Language Universals*. PhD thesis, University of Edinburgh.
- Kirby, S. (1999). *Function, Selection and Innateness: The emergence of language universals*. Oxford University Press.
- Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Science*, 104:5241–5245.
- Kirchner, R. (1998). *An Effort-Based Approach to Consonant Lenition*. PhD thesis, UCLA.
- Kirchner, R. and Moore, R. (2009). Computing phonological generalization over real speech exemplars. ms.
- Klein, S. (1966). Historical change in language using Monte Carlo techniques. *Mechanical Translation and Computational Linguistics*, 9.

- Klein, S., Kuppin, M., and Meives, K. (1969). Monte Carlo simulation of language change in Tikopia and Maori. In *Proceedings of the 1969 Conference on Computational Linguistics*, pages 699–729.
- Kornai, A. (1990). Hungarian vowel harmony. In *Approaches to Hungarian, Volume Three: Structures and Arguments*. Szeged.
- Krämer, M. (2003). *Vowel Harmony and Correspondence Theory*, volume 66 of *Studies in Generative Grammar*. Walter de Gruyter.
- Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1):22–44.
- Labov, W. (1994). *Principles of Linguistic Change, Vol. 1: Internal factors*. Blackwell.
- Langacker, R. (2008). *Cognitive Grammar: A basic introduction*. Oxford University Press.
- Lass, R. (1997). *Historical Linguistics and Language Change*. Cambridge University Press.
- Leiwo, M., Kulju, P., and Aoyama, K. (2006). The Acquisition of Finnish Vowel Harmony. In *A Man of Measure: Festschrift in Honour of Fred Karlsson*, pages 149–161. The Linguistic Association of Finland. Special supplement to Vol. 19 of SKY Journal of Linguistics.
- Lightner, T. (1965). On the description of vowel and consonant harmony.

*Word*, 21:224–250.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35:1773–1781.

Lindblom, B. (1990). *Explaining phonetic variation: a sketch of the H and H theory.*, pages 403–439. Kluwer Academic Publishers.

Lodge, K. (2009). *Fundamental Concepts in Phonology: Sameness and difference.* Edinburgh University Press.

MacKay, D. J. (2002). *Information Theory, Inference and Learning Algorithms.* Cambridge University Press.

MacWhinney, B. (1978). *The Acquisition of Morphophonology.* Monographs of the Society for Research in Child Development. Blackwell.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 25:187–205.

Mahanta, S. (2007). *Directionality and Locality in Vowel Harmony.* PhD thesis, Utrecht University.

Mailhot, F. and Reiss, C. (2007). Computing Long-distance Dependencies in Vowel Harmony. *Biolinguistics*, 1:28–48.

Manning, C. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing.* MIT Press.

Manuel, S. (1999). Cross-language studies: relating language-particular

- coarticulation patterns to other language-particular facts. In Hardcastle, W. and Hewlett, N., editors, *Coarticulation: Theory, data and techniques*, pages 179–198. Cambridge University Press.
- Marcus, G. (1993). Negative evidence in language acquisition. *Cognition*, 46:53–85.
- Marcus, G. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Learning, Development, and Conceptual Change. MIT Press.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.
- Medin, D. and Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, 85:207–238.
- Mielke, J. (2008). *The Emergence of Distinctive Features*. Oxford Studies in Typology and Linguistic Theory. Oxford University Press.
- Mielke, J. and Lin, Y. (2007). Discovering place and manner features - What can be learned from acoustic and articulatory data? In *Proceedings of the 31st Penn Linguistics Colloquium*.
- Miller, G. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, 7:141–144.
- Mitchell, T. (1997). *Machine Learning*. McGraw Hill.

- Morais, J., Cary, L., Alegria, J., and Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7:323–331.
- Murphy, G. (2002). *The Big Book of Concepts*. MIT Press.
- Neisser, U. (1967). *Cognitive Psychology*. Prentice Hall.
- Niyogi, P. (2006). *The Computational Nature of Language Learning and Evolution*. MIT press.
- Niyogi, P. and Berwick, R. (1998). The logical problem of language change: A case study of European Portuguese. *Syntax*, 1:192–205.
- Noske, M. (2000). [ATR] harmony in Turkana : A case of Faith(suffix) >> Faith(root). *Natural Language and Linguistic Theory*, 18(4):771–812.
- Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115:39–57.
- Ohala, J. J. (1981). The listener as a source of sound change. In Masek, C., Hendrick, R., and Miller, M., editors, *Papers from the Parasession on Language and Behavior*, pages 178–203.
- Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In Breivik, L. and Jahr, E., editors, *Language Change: Contributions to the study of its causes.*, pages 173–198. Mouton de Gruyter.

- Ohala, J. J. (1993). Coarticulation and Phonology. *Language and Speech*, 36:155–170.
- Ohala, J. J. (1994a). Hierarchies of environments for sound variation; plus implications for ‘neutral’ vowels in vowel harmony. *Acta Linguistica Hafniensa*, 27:371–382.
- Ohala, J. J. (1994b). Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of the Third International Conference on Spoken Language Processing*, pages 491–494, Yokohama, Japan.
- Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39:151–168.
- Oliphant, T., Jones, E., Peterson, P., et al. (2001). Scipy: Open source scientific tools for python.
- Oudeyer, P.-Y. (2006). *Self-Organization in the Evolution of Speech*. Studies In The Evolution of Language. Oxford University Press.
- Phillips, C. (2001). Levels of representations in the electrophysiology of speech perception. *Cognitive Science*, 25:711–731.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J. and Hopper, P., editors, *Frequency effects and the emergence of linguistic structure*, pages 137–157. John Benjamins.
- Pierrehumbert, J. (2003). Probabilistic Phonology: Discriminations and

- Robustness. In Bod, R., Hay, J., and Jannedy, S., editors, *Probabilistic Linguistics*, pages 177–228. MIT Press.
- Port, R. (2007a). The graphical basis of phones and phonemes. In Bohn, O.-S. and Munro, M., editors, *Language Experience in Second Language Speech Learning*, pages 349–365. John Benjamins.
- Port, R. (2007b). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, 25:143–170.
- Port, R. and Leary, A. (2005). Against formal phonology. *Language*, 81:927–964.
- Prince, A. and Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Blackwell.
- Pulleyblank, D. (1996). Neutral Vowels in Optimality Theory: A comparison of Yoruba and Wolof. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 41:295–347.
- Purnell, T., Raimy, E., and Salmons, J. (2010). Bio-cognitive modularity and sound systems. Presented at 2010 LSA Annual Meeting.
- Rabiner, L. and Juang, B.-H. (1993). *Fundamentals of Speech Recognition*. Prentice Hall.
- Raimy, E. (2000). *The phonology and morphology of reduplication*. Mouton de Gruyter.

- Ratner, N. B. and Menn, L. (2000). In the beginning was the wug: Forty years of language elicitation studies. In *Methods for Studying Language Production*. Lawrence Erlbaum.
- Read, C., Yun-Fei, Z., Hong-Yin, N., and Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, 24:31–44.
- Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, 76:1624–1635.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4:328–350.
- Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, 104:192–233.
- Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1 edition.
- Salmon, W. (1998). *Causality and Explanation*. Oxford University Press.
- Samuels, B. (2009). *The Structure of Phonological Theory*. PhD thesis, Harvard University, Cambridge, MA.
- Schölkopf, B. and Smola, A. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.
- Semon, R. (1921). *The Mneme*. George Allen and Unwin.



- Silverman, D. (2006). *A Critical Introduction to Phonology: Of Sound, Mind, and Body*. Continuum Critical Introductions to Linguistics. Continuum.
- Skousen, R. (1989). *Analogical Modeling of Language*. Kluwer Academic Publishers.
- Skousen, R., Lonsdale, D., and Parkinson, D., editors (2002). *Analogical Modeling: An exemplar-based approach to language*, volume 10 of *Human Cognitive Processing*. John Benjamins.
- Smolensky, P. (1993). Harmony, markedness, and phonological activity. Presented at Rutgers Optimality Workshop 1. Rutgers University: New Brunswick, NJ.
- Steels, L. (1997). The synthetic modeling of language origins. In *Evolution of Communication*. John Benjamins.
- Stevens, K. (2000). *Acoustic Phonetics*. MIT Press.
- Stevens, K. (2002). Toward a model of lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111:1872–1891.
- Stevens, K. and Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In Wathen-Dunn, W., editor, *Models for the Perception of Speech and Visual Form.*, pages 88–102. MIT Press.
- Sussman, H., Fruchter, D., Hilbert, J., and Sirosh, J. (1998). Linear correlates

- in the speech signal: The orderly output constraint. *Behavioral and Brain Sciences*, 21:241–299.
- Tulving, E. (1972). Episodic and semantic memory. In Tulving, E. and Donaldson, W., editors, *Organization of memory*, pages 381–403. Academic Press.
- Välimaa-Blum, R. (2009). The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? *Cognitextes*, 2. Retrieved from <http://cognitextes.revues.org/211> on 2010-07-16.
- van der Hulst, H. and van de Weijer, J. (1995). Vowel harmony. In Goldsmith, J., editor, *Handbook of Phonological Theory*. Blackwell.
- Wedel, A. (2004a). Category competition drives contrast maintenance within an exemplar-based production/perception loop. In Goldsmith, J. and Wicentowski, R., editors, *Proceedings of the Seventh Meeting of the ACL Special Interest Group in Computational Phonology*, pages 1–10.
- Wedel, A. (2004b). *Self-organization and Categorical Behavior in Phonology*. PhD thesis, UC Santa Cruz.
- Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, 23:247–274.
- Wedel, A. (2007). Feedback and regularity in the lexicon. *Phonology*, 24:147–185.

- Weinreich, U., Labov, W., and Herzog, M. (1968). Empirical foundations for a theory of language change. In Lehmann, W. and Malkiel, Y., editors, *Directions for historical linguistics*, pages 95–195. University of Texas Press.
- Wheeler, D. and Touretzky, D. (1993). A connectionist implementation of cognitive phonology. In *The Last Phonological Rule*, pages 1446–172. University of Chicago Press.
- Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford University Press.
- Zimmer, K. (1985). Arabic loanwords and turkish phonological structure. *International Journal of American Linguistics*, 51:623–625.