Instance-based acquisition of vowel harmony

Frédéric Mailhot
Institute of Cognitive Science
Carleton University
Ottawa, ON, Canada
`fmailhot@connect.carleton.ca`

Abstract

I present LIBPHON, a nonparametric regression-based model of phonological acquisition that induces a generalised and productive pattern of vowel harmony—including opaque and transparent neutrality—on the basis of simplified formant data. The model quickly learns to generate harmonically correct morphologically complex forms to which it has not been exposed.

1 Explaining phonological patterns

How do infants learn the phonetic categories and phonotactic patterns of their native languages? How strong are the biases that learners bring to the task of phonological acquisition? Phonologists from the rationalist tradition that dominated the past half-century of linguistic research typically posit strong biases in acquisition, with language learners using innately-given, domain-specific representations (Chomsky and Halle, 1968), constraints (Prince and Smolensky, 2004) and learning algorithms (Tesar and Smolensky, 2000; Dresher, 1999) to learn abstract rules or constraint rankings from which they can classify or produce novel instances.

In the last decade, however, there has been a shift toward empiricist approaches to phonological acquisition, use and knowledge. In this literature, *eager* learning algorithms (Aha, 1997), in which training data are used to update intensional representations of functions or categories then discarded, have been the norm.[1] However, research in related fields—particularly speech perception—indicates that speakers' knowledge and use of language, both in production and comprehension, is at least partly episodic, or instance-based (Goldinger, 1996; Johnson, 1997). Additionally,

motivation for instance-based models of categorisation has a lengthy history in cognitive psychology (Medin and Schaffer, 1978), and these methods are well-known in the statistical and machine learning literature, having been studied for over half a century (Fix and Hodges, 1951; Cover and Hart, 1967; Hastie et al., 2009). Consequently, it seems a worthy endeavour applying an instance-based method to a problem that is of interest to traditional phonologists, the acquisition and use of vowel harmony, while simultaneously effecting a *rapprochement* with adjacent discplines in the cognitive sciences. In sections 2 and 3 I give some brief background on vowel harmony and instance-based models, respectively. Section 4 introduces my model, LIBPHON, and section 5 the languages it learns. I discuss some simulations and results in section 6, and conclude in section 7.

2 Vowel harmony

Vowel harmony is a phonological phenomenon in which there are co-occurrence constraints on vowels within words.[2] The vowels in a language with vowel harmony can be classified into disjoint sets, such that words contain vowels from only one of the sets. The Finnish system of vowel harmony exemplified by the forms in Table 1 provides a standard example from the literature (van der Hulst and van de Weijer, 1995).

	surface form	*gloss*
a.	tuhmasta	'naughty' (elative)
b.	tühmästä	'stupid' (elative)

Table 1: Finnish backness harmony

Crucially, the elative case marker alternates systematically between front and back vowel

---

[1]Daelemans et al. (1994) is a notable exception.

[2]"Word" is used pre-theoretically here; harmony can occur over both supra- and sublexical domains.

variants—as *-stä* or *-sta*—depending on whether the stem has front $\{\ddot{u}, \ddot{a}\}$ or back $\{u, a\}$ vowels.

2.1 Neutral vowels

In most languages with vowel harmony, there are one or more vowels that systematically fail to alternate. These are called *neutral vowels*, and are typically further subclassified according to whether or not they induce further harmonic alternations in other vowels.

3 Instance-based models

Instance-based approaches to cognitive processing, also called memory-based, case-based, and exemplar-based models, have their modern origins in psychological theories and models of perceptual categorisation and episodic memory (Medin and Schaffer, 1978; Nosofsky, 1986), although the earliest explicit discussion seems to be (Semon, 1921); a theory of memory that anticipates many features of contemporary models. The core features of these models are: (i) explicit storage/memorisation (*viz.* extensional representation) of training data, (ii) classification/processing of novel data via similarity-based computation, and (iii) *lazy evaluation* (Aha, 1997), whereby all computations are deferred until the model is queried with data.[3]

Instance-based models were introduced to linguistics via research in speech perception suggesting that at least some aspects of linguistic performance rely on remembered experiential episodes (Johnson and Mullenix, 1997). The models implemented to date in phonetics and phonology have largely focused on perception (*e.g.* speaker normalisation in Johnson (1997)), or on diachronic processes (*e.g.* lenition in Pierrehumbert (2001), chain shifts in Ettlinger (2007)), leaving the types of phenomena that typically interest "traditional" phonologists, *viz.* productive, generalised patterns, comparatively neglected.[4]

4 LIBPHON

LIBPHON, the **L**azy **I**nstance-**b**ased **Phon**ologist, is a lazy learning algorithm whose purpose (in the context of the simulations described here) is to model an instance-based approach to the core aspects of the acquisition and subsequent productive usage of vowel harmony.

4.1 Decisions & mechanisms

As discussed in (Johnson, 2007), there are some decisions that need to be made in implementing an instance-based model of phonological knowledge involving the basic units of analysis (*e.g.* their size), the relevant type of these units (*e.g.* discrete or continuous), and the mechanisms for similarity-matching and activation spread in the lexicon.

Units The arguments given by Johnson (2007) and Välimaa-Blum (2009) for the "word-sized" (rather than *e.g.* segmental) experience of language, suggest that "words" are the correct basic unit of analysis in instance-based langugage models (*a fortiori* in LIBPHON). Stronger evidence comes from the wealth of psycholinguistic data (reviewed in (Lodge, 2009)) showing that illiterates and literates of non-alphabetic writing systems have poor phonemic (or at least segmental) awareness, both in monitoring and manipulation. On this basis, I take *meaning-bearing unanalysed acoustic chunks* to be the relevant units of representation for LIBPHON.[5]

Feature type Having determined the size of LIBPHON's basic unit, I move now to its embedding space, where distinctive features present themselves as obvious candidate dimensions. Since the middle of the 20th century (*ca.* Chomsky and Halle (1968)), phonological theories have nearly all supposed that lexical representations are stored in terms of articulatory features (*cf.* (Halle, 1997) for explicit discussion of this viewpoint). Coleman (1998), citing evidence from the neuroscientific and psycholinguistic literatures on lexical representation, claims that evidence for this position (*e.g.* from speech perception and phoneme monitoring experiments) is weak at best, and that lexical representations are more likely to be acoustic than articulatory. In addition, Phillips et al. (2000) review neurolinguistic evidence for the role of acoustic cortex in phonetics and phonology, and

---

[3]Compare *eager* learners, *e.g.* connectionist systems, which build a global intensional representation of the function being learned on the basis of training data which are subsequently discarded.

[4]Kirchner and Moore (2009) give a model of a synchronic lenition process, and Daelemans and colleagues give memory-based analyses of several linguistic phenomena (Daelemans and van den Bosch, 2005).

[5]The assumption that word-level segmentation of the speech signal is available to the language learner prior to acquisition of phonological phenomena is relatively uncontroversial, although there is evidence for the development of at least some phonotactic knowledge prior to the emergence of a productive lexicon (Jusczyk, 1999).

Mielke (2008) discusses several aspects of the induction of distinctive phonological features from acoustic representations. Recognising that the issue is far from resolved, for the purposes of the simulations run here, I take LIBPHON's instance space to be acoustically-based, and use formant values as the embedding dimension. Vowels are specified by their midpoint formant values,[6] and consonants are specified by so-called "locus" values, which can be identified by inspecting the trajectories of consonant-vowel transitions in speech (Sussman et al., 1998). Since I am modelling palatal harmony in particular, and $F2$ magnitude is the primary acoustic correlate of vowel palatality, I omit $F3$ and $F4$, restricting LIBPHON's acoustic representations to sequences of $(F1, F2)$ values, henceforth *trajectories*.

Similarity Given that LIBPHON's instance-space is continuous, and has a fairly intuitive metric, I take simple Euclidean distance to be LIBPHON's similarity (or rather, dissimilarity) function.[7]

Fixed-rate representations For the simulations described here, I use fixed-rate trajectories, in which consonants and vowels are represented in a temporally coarse-grained manner with single $(F1, F2)$ tuples. Evidently, consonants and vowels in actual human speech unfold in time, but modelling segments 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 main obstacles in the development of instance-based models of speech production, due to the difficulty of aligning variable-length forms. Although algorithms exist for aligning variable-length sequences, these require cognitively implausible dynamic programming algorithms, *e.g.* dynamic time warping (DTW)

and hidden Markov models (Rabiner and Juang, 1993). Even as proofs of concept, these may be empirically inadequate; Kirchner and Moore (2009) use DTW to good effect in an instance-based production model of spirantisation using real, temporally variable, speech signals. However, their inputs were all the same length in terms of segmental content, and the model was only required to generalise within a word type. I am currently investigating whether DTW can function as a proof of concept in a problem domain like that addressed here, which involves learning about variably-sized "pieces" of morphology across class labels.

4.2 Perception/categorisation

LIBPHON's method of perception/categorisation of inputs is a relatively standard nearest-neighbour-based classification algorithm. See Algorithm 1 for a description in pseudocode.

Algorithm 1 PERCEIVE(*input*, k)

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

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

If LABEL is not empty, LIBPHON checks its lexicon to see whether it knows the word being presented to it, *i.e.* whether it exists as a class label. If so, it simply appends the input acoustic form to the set of forms associated with the input meaning/label. If it has no corresponding entry, a new lexical entry is created for the input meaning, and the input trajectory is added as its sole associated acoustic form.

If LABEL is empty, LIBPHON assigns `instance` to the majority class of its k nearest neighbours in acoustic space.

---

[6]A reviewer asks about the psychological plausibility of Hz-based formant represetations and the choice of point values for vowel and consonant representations, *e.g.* rather than formant values at 20% and 80% of the vowel. These are purely in the interests of simplicity for the work reported here. As discussed below, future work with real speech exemplars in psychophysically-motivated representational formats, *e.g.* perceptual linear predictive coding (Hermansky, 1990), will render this issue moot.

[7]Often the measure of similarity in an instance-based model is an exponential function of distance, $d(x_i, x_j)$ of the form $exp(-cd(x_i, x_j))$, so that increasing distance yields decreasing similarity (Nosofsky, 1986). The Euclidean measure here is sufficient for the purpose at hand, although the shape of the similarity measure is ultimately an empirical question.

4.3 Production

In production, LIBPHON is provided with a LA-BEL and has to generate a suitable `instance` for it. LABELs are decomposable, signalling an arbitrary "lexical" meaning, an optional plural morpheme, PL, and an obligatory case marker from {NOM, ACC}. Thus, there are several different possibilities to consider in generating output for some queried meaning.

In the two simplest cases, either the full queried meaning (*viz.* lexical label with all inflections) is already in the lexicon, or else there are no class LABELs with the same lexical meaning (*i.e.* LIBPHON is being asked to produce a word that it doesn't know). In the former case, a stored trajectory is uniform[8] randomly selected from the list of acoustic forms associated with the queried label as a seed token, the entire set of associated acoustic forms is used as the analogical set, and an output is generated by taking a distance-weighted mean over the seed's k nearest neighbours.[9] In the case where the lexical meaning of the queried LABEL is unknown, the query is ignored.

In the more interesting cases, LIBPHON has a LABEL in its lexicon with the same lexical meaning, but with differing inflectional specification. Consider the case in which LIBPHON knows only the singular NOM form of a query label that is specified as PL ACC. A seed `instance` is (uniform) 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. In this case the analogical set, the set of instances from which the final output is computed, is composed of the seed's nearest neighbours in the set of all trajectories associated with LABELs of the form [LEX PL ACC]. Once 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 LA-BEL 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 LA-BEL
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

4.4 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 in particular are closely related to the well-known Nadaraya-Watson estimator (Hastie et al., 2009):

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

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

---

[8] Exemplar models often bias the selection of seed tokens with degrees of "activation" that take into account recency and frequency. Although the results discussed below show that this is not necessary for ultimate attainment, it is likely that this kind of bias will need to be incorporated into LIBPHON to accurately model more nuanced aspects of the acquisition path.

[9] $k = 5$ for all results reported here.

5 The languages

On the view taken here, phonological knowledge is taken to emerge from generalisation over lexical items, and so the key to acquiring some phonological pattern lies in learning a lexicon (Jusczyk, 2000). Consequently, the languages learned in LIBPHON abstract away from sentence-level phenomena, and the training data are simply labelled formant trajectories, (LABEL, `instance`).

In order to get at the essence of the problem (*viz.* the acquisition of vowel harmony as characterised by morphophonological alternations), and in the interests of computational tractability/efficiency, the artificial languages learned by LIBPHON are highly simplified, displaying only enough structure to capture the phenomena of interest.

5.1 Phonological inventory

The phonological inventory consists of three consonants, $\{b, d, g\}$, and four vowels—two with high $F2$ and two with low $F2$—which I label $\{i, e, u, o\}$, for convenience.[10] The formant values used were generated from formant synthesis equations in (de Boer, 2000), and from the locus equations for CV-transitions in (Sussman et al., 1998).

5.2 Lexical items

LIBPHON's lexicon is populated with `instance` trajectories consisting of four-syllable[11] "roots" with zero, one or two one-syllable "affixes". These trajectories have associated class labels, which from a formal point of view are contentless indices. Rather than employing *e.g.* natural numbers as labels, I use character strings which correspond more or less to the English pronunciations of their associated trajectories. LABELs function, metaphorically-speaking, as "meanings". These are compositional, comprising a "lexical meaning" (arbitrary CVCVCVCV string from the phoneme set listed above), one of two obligatorily present "case markers" (NOM|ACC), and an optionally present "plural marker" (PL). Hence, word categories in the artificial languages come in four forms, NOM-SG, NOM-PL, ACC-SG, and ACC-PL.
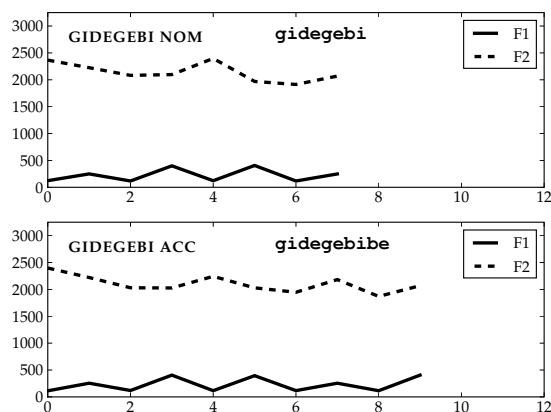

Figure 1: Graphical representation of singular forms of GIDEGEBI, as produced by teacher agent

Figure 1 gives examples of the singular NOM and ACC forms of a high-$F2$ word. The NOM-labelled trajectory has no suffixal morphology, and corresponds to a bare form. The trajectory is eight segments long,[12] and the vowels in this case have all high $F2$ (as in lexical front/back vowel harmony).[13] Note also that ACC is realised with high $F2$, in agreement with the root vowels.

5.3 Neutral vowels

The harmony processes seen thus far are in some sense "local", being describable in terms of vowel adjacency *e.g.* adjacency on a hypothesised autosegmental tier (although the presence of intervening consonants still renders the harmony process "nonlocal" in some more concrete articulatory sense). One of the hallmarks of vowel harmony, as discussed in subsection 2.1, is the phenomenon of *neutral vowels*. These vowels fail to alternate, and may or may not induce harmonic alternations in vowels that precede or follow them. 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.

Figures 2 and 3 show the "plural inflected" forms of the same root as in 1. We see that the

---

[10]Because LIBPHON's representations lack $F3$, the primary acoustic correlate of rounding, the back vowels would be more standardly represented as /ɤ, ɯ/, but these are comparatively uncommon IPA symbols, so we will use the symbols for the rounded variants. Nothing in the results or discussion hinges on this.

[11]All syllables are CV-shaped.

[12]The even-numbered indices on the x-axis correspond to consonants and the odd-numbered indices, the "pinches" in the graphs, correspond to vowels.

[13]The languages LIBPHON learns have only harmonic singular forms. This is unrealistic, as speakers of vowel harmony languages typically have some exceptional disharmonic forms in their lexicons. The effect of these forms on LIBPHON's performance is currently being investigated.

realisation of PL has fixed, low $F2$, and that the realisation of ACC has alternating $F2$, which realisations corresponding roughly to [be] (high $F2$) and [bo] (low $F2$).
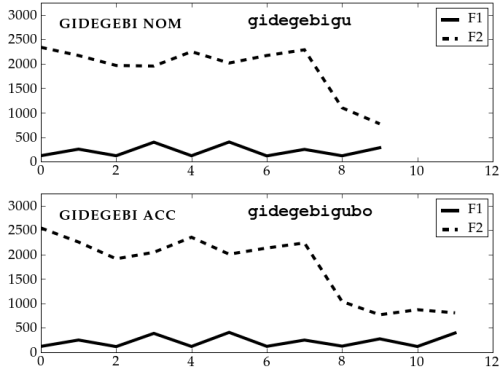

Figure 2: Graphical representation of plural forms of GIDEGEBI, as produced by teacher agent, with opaque PL realisation.
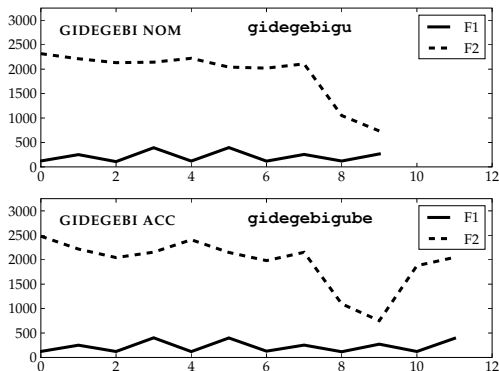

Figure 3: Graphical representation of plural forms of GIDEGEBI, as produced by teacher agent, with transparent PL realisation.

These figures also illustrate the difference between languages with opaque versus transparent PL, as reflected in the realisation of the word-final ACC marker in the two lower graphs, which agrees in $F2$ with the realised form of the PL or root, respectively.

6 The experiments

Assessing successful learning/generalisation in a computational model requires some measurable outcome that can be tracked over time. Because LIBPHON is an output-oriented model, its cate-

gorisation of inputs is a poor indicator of the extent to which it has learned a productive "rule" of vowel harmony. In lieu of this measure, I have opted to pursue two difference courses of evaluation.

For the harmony cases, LIBPHON is queried on a held-out test set of 500 previously unseen LABELs and its output is compared to the mean value of the teacher's stored trajectories for the same LABELs. In particular, given some LABEL which was not in the training data, we can query LIBPHON at various stages of acquisition (*viz.* with lexicons of increasing size) by having it produce an output for that LABEL, and track the change in its performance over time.

The actual measure of error taken is the root-mean-squared deviation between the learner's output, y and the mean, t, of the teacher's stored forms for some label, l, over all of the consonants and vowels within a word, averaged across the remaining unseen items of the test set:

$$RMSE = \frac{1}{N} \sum_{l \in lex} \sqrt{\frac{\sum_i (t_i - y_i)^2}{len(t)}}$$

Figures 4 and 5 show RMSE *vs.* lexicon size for both opaque and transparent neutrality (*cf.* the cases in Figures 2 and 3), for five simulation runs each. We can see clearly that error drops as the lexicon grows, hence that LIBPHON is learning to make its outputs more like those of the teacher, but the informativity of this measure stops there. From a linguistic point of view, we are interested in what LIBPHON's outputs look like, *viz.* has it learned vowel harmony?
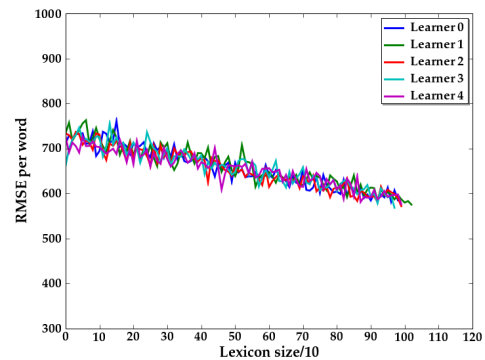

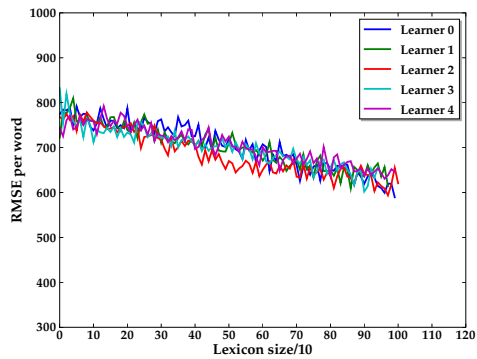Figure 4: RMSE 1000-word lexicon. Opaque neutrality.

Figure 5: RMSE 1000-word lexicon. Transparent neutrality.

Figure 6: Evolution of `gubogobu` in early acquisition: 10 words

Figure 7: Evolution of `gubogobu` in early acquisition: 30 words

Figures 6 and 7 show that vowel harmony *is* learned, and moreover quite quickly, after going through a brief initial phase of spurious outputs. In these figures, LIBPHON is being asked to produce outputs for all forms of the label GUBOGOBU. For the particular run shown here, at the 10-word stage (*i.e.* when LIBPHON had seen tokens from 10 labels), the only tokens marked PL-ACC were from high $F2$ ("front") trajectories. Hence the nearest neighbour calculation in the production algorithm resulted in a fronted form being output. Although acquisition research in vowel harmony languages is relatively rare, or inaccessible to us due to language barriers, what research there is seems to indicate that harmony is mastered very quickly, with virtually no errors by 2 years of age, hence it is unclear what status to assign to output patterns like the one discussed here. Moreover, given the well-known facts that (i) comprehension precedes production, and (ii) infants avoid saying unfamiliar words, it is unlikely that an infant could be coaxed into producing an output form for such an early-stage class.

7 Discussion and future work

The experiments discussed here show that on the basis of limited input data, LIBPHON, an instance-based learner that produces output via kernel-smoothed nearest-neighbour regression, learns to produce harmonically correct novel outputs. In particular, it is able to generalise and produce correct morphologically complex forms to which it has not been exposed in its training data, *i.e.* a previously unseen case-marked form will be output with harmonically correct $F2$,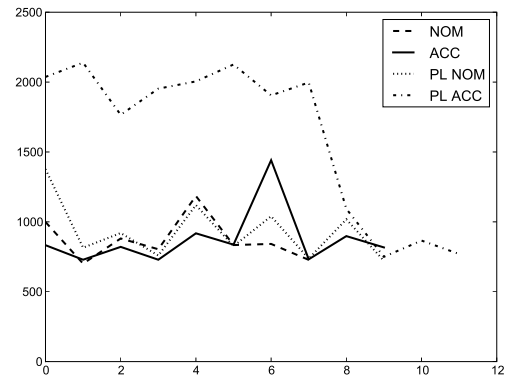 including neutrality (opaque or transparent). In ongoing research I am (i) evaluating LIBPHON's performance with respect to more traditional measures, in particular F-score, on held-out data as the lexicon grows, and (ii) assessing the viability of DTW-based alignment for preprocessing real speech tokens as inputs to LIBPHON.

References

David Aha. 1997. Lazy learning. In *Lazy Learning*, pages 7–10. Kluwer Academic Publishers.

7

Noam Chomsky and Morris Halle. 1968. *The Sound Pattern of English*. Harper and Row.

John Coleman. 1998. Cognitive reality and the phonological lexicon: A review. *Journal of Neurolinguistics*, 11(3):295—-320.

Thomas Cover and Peter Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27.

Walter Daelemans and Antal van den Bosch. 2005. *Memory-Based Language Processing*. Studies in Natural Language Processing. Cambridge University Press.

Walter Daelemans, Steven Gillis, and Gert Durieux. 1994. The acquisition of stress: A data-oriented approach. *Computational Linguistics*, 20(3).

Bart de Boer. 2000. Self-organization in vowel systems. *Journal of Phonetics*, 28:441–465.

B. Elan Dresher. 1999. Charing the learning path: Cues to parameter setting. *Linguistic Inquiry*, 30(1):27–67.

Marc Ettlinger. 2007. An exemplar-based model of chain shifts. In *Proceedings of the 16th International Congress of the Phonetic Science*, pages 685–688.

Evelyn Fix and J.L. Hodges. 1951. Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine.

Stephen Goldinger. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1166–1183.

Morris Halle. 1997. Some consequences of the representation of words in memory. *Lingua*, 100:91–100.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *Elements of Statistical Learning*. Springer Series in Statistics. Springer-Verlag, 2 edition.

Hynek Hermansky. 1990. Perceptual linear predictive (plp) analysis of speech. *Journal of the Acoustical Society of America*, 87(4):1738–1752.

Keith Johnson and John W. Mullenix, editors. 1997. *Talker Variability in Speech Processing*. Academic Press.

Keith Johnson. 1997. Speech perception without speaker normalization: an exemplar model. In *Talker Variability in Speech Processing*, chapter 8, pages 145–166. Academic Press.

Keith Johnson, 2007. *Decision and Mechanisms in Exemplar-based Phonology*, chapter 3, pages 25–40. Oxford University Press.

Peter Jusczyk. 1999. How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3(9):323–328.

Peter Jusczyk. 2000. *The Discovery of Spoken Language*. MIT Press.

Robert Kirchner and Roger Moore. 2009. Computing phonological generalization over real speech exemplars. ms.

Ken Lodge. 2009. *Fundamental Concepts in Phonology: Sameness and difference*. Edinburgh University Press.

Douglas Medin and Marguerite Schaffer. 1978. Context theory of classification learning. *Psychological Review*, 85(3):207–238.

Jeff Mielke. 2008. *The Emergence of Distinctive Features*. Oxford Studies in Typology and Linguistic Theory. Oxford University Press.

Robert Nosofsky. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1):39–57.

Colin Phillips, Thomas Pellathy, Alec Marantz, Elron Yellin, Kenneth Wexler, David Poeppel, Martha McGinnis, and Timothy Roberts. 2000. Auditory cortex accesses phonological categories: An meg mismatch study. *Journal of Cognitive Neuroscience*, 12(6):1038–1055.

Janet Pierrehumbert. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency effects and the emergence of linguistic structure*, pages 137–157. John Benjamins.

Alan Prince and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Blackwell.

Lawrence Rabiner and Biing-Hwang Juang. 1993. *Fundamentals of Speech Recognition*. Prentice Hall.

Richard Semon. 1921. *The Mneme*. George Allen and Unwin.

Harvey Sussman, David Fruchter, Jon Hilbert, and Joseph Sirosh. 1998. Linear correlates in the speech signal: The orderly output constraint. *Behavioral and Brain Sciences*, 21:241–299.

Bruce Tesar and Paul Smolensky. 2000. *Learnability in Optimality Theory*. MIT Press.

Riitta Välimaa-Blum. 2009. The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? *Cognitextes*, 2.

Harry van der Hulst and Jeroen van de Weijer. 1995. Vowel harmony. In John Goldsmith, editor, *Handbook of Phonological Theory*. Blackwell.