Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 34 (2006) 516–530
Journal of Phonetics
www.elsevier.com/locate/phonetics

The next toolkit

Janet B. Pierrehumbert*

Department of Linguistics & Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60201, USA

Received 10 April 2006; received in revised form 28 May 2006; accepted 8 June 2006

1. Introduction

Language is an example of collective behavior, and it is a type of collective behavior for which people are highly adapted. Comparable systems in other creatures in no way approach the level of organization found in human languages. Language provides a canonical example of a complex system: robust, adaptable, and self-assembling. Self-assembly occurs both in the cognitive system (as children bootstrap a complex generative system from limited and varied exposure to its manifestations) and in populations, as people match their language systems to each other, and group themselves into social networks of people who share the same language.

Linguistic systems arise within the affordances of physics, biology, and society. Take vowels as an example. The fact that vowels can be well characterized using three formant trajectories (disregarding all resonances of the vocal tract above the third one) derives from several facts. One is that the glottal source spectrum rolls off rapidly, with the result that upper formants are not reliably excited. Another is that the hearing systems of people and all other animals effectively compute a spectral representation of the sound, from which the resonances of the sound source can be estimated. Yet another is that the auditory system is most sensitive right in the region of the first three formants. The confluence of these factors as the phonetic grounding of the vowel map is discussed at more length in Pierrehumbert (2000).

Not all combinations of three resonances correspond to possible vowels, of course; the entire three-dimensional vowel space is effectively circumscribed by the articulatory range of human beings. Thus, physics and biology define a vowel solid, much as the color vision system defines a color solid. Specific languages preferentially exploit different regions of this solid. For example, the region corresponding to cardinal [u] is not exploited in many dialects of American English. In the American South, the closest correspondant to [u] is much fronter; Labov, Ash, and Boberg (2006) report F2 values in the range of 1400–2000 Hz for this region. Fronter than cardinal [u], but not as front as French /y/, the Southern AE [u] corresponds to a particular region in the three-dimensional vowel space, much as the color word "teal" corresponds to a particular region of the color solid which may not have a precise label in some other languages.

Now, we can also consider the cluster of people who share any given variant of /u/. In a dialect atlas, such as Labov et al. (2006), the areas in which a variant predominates are marked as regions on a map. Of course this

*Tel.: +1 847 491 5779; fax: +1 847 491 3770.
E-mail address: jbp@northwestern.edu.

doi:10.1016/j.wocn.2006.06.003

is a simplification. Considered in more detail, the human population is a network in which each node is a person, and a (phonological) dialect is a clustering of people with consistent phonological systems. Such people are disproportionately, though not absolutely, congregated in space. The congregation is not absolute because the actual proximity measure is social affiliation, which is only partially correlated with spatial proximity.

Social affiliation matters because phonological systems become consistent through social interaction. Psycholinguistic literature reviewed in Locke (1993) shows the role of social bonding in language acquisition from its most incipient stages in the newborn infant. Preschool children match the dialectal features of their primary caregivers, and modify these features when they start school and acquire a peer group of schoolmates. Implicit choices in speech patterns continue to be made during adolescence, with phonetic details serving as markers of social identity. And evidence for continuing malleability, in accommodation to changes in the speech community in adult life, has now begun to trickle in, most notably through the work of Harrington and colleagues (this issue and references cited there).

The basic mechanism by which social interaction induces consistency in linguistic systems was already sketched in Wiener (1948) and in Denes and Pinson's (1963) classic high school textbook, *The Speech Chain* (2nd edition, 1963). It is a feedback loop through interlocutors in the speech community. Each listener perceives and categorizes the speech produced by others. When it comes to be their turn to speak, they adopt the familiar patterns as production goals. A variety of models have shown that entrainment occurs in models in which speakers correct errors and/or have independent knowledge of the intended meaning (De Boer, 2000; Wedel, 2004). Oudeyer (2006) and Pierrehumbert (2004) have shown that entrainment in speech patterns amongst interlocutors can occur even in the absence of any semantic content or error correction, A notable feature of the Oudeyer and Pierrehumbert models is that speakers do not in any way calculate the listener's needs, responding to experimental findings showing that speakers are really not very good at such calculations (Bard et al., 2000). Entrainment amongst speakers results simply from the phonetic categorization and the entrenchment of the perception-production loop through reuse. Furthermore, all models of this class that include a parametric phonetic map yield outcomes in which fine phonetic details are learned from experience, and can therefore differ from one language or dialect to another.

Given these results of computational modeling, we can understand systems of phonological categories as equilibria (or rather, near-equilibria) at the population level. The underlying physical and biological systems support many such equilibria, in fact very many; there is no known example of two languages with literally identical phonological category systems. The interaction of random variation with the circular causality of population-level feedback explains how such detailed differences can come about against a backdrop of common biological and cognitive endowments. If we add the further, obvious, assumption that people select their social contacts in large part on the basis of common language, it is clear that language system and the social system are interdependent. Language affects the social system, and the social system affects the language. There are no chicken-and-egg problems in this approach; the locus of explanation lies in the exact nature of the interdependencies.

The innovations of Goldinger (1996, 1998) and Johnson (1997) provide important new tools for formalizing this general picture, and these papers serve as an explicit or implicit backdrop to all of the papers in this volume. These authors imported exemplar theory into the speech processing community, where it was subsequently taken up by Pierrehumbert (2001), and many others. They proposed that cognitive representations of speech are extremely detailed; people can implicitly learn phonetic distributions associated with speech segments and even with individual words. Richly indexed storage of experiences means that the frequencies of different factors and their density distributions (whether Gaussian or not) can be represented. These proposals provide a way to theoretically instantiate the lessons of many years of work in linguistic typology, which showed that parametric distributions are so language- and context-particular that they must surely be learned. Goldinger (1996, 1998) proposed a mechanism for continual updating of the density distributions in the speech processing system. Such continual updating can capture otherwise puzzling interactions between the nature of experiences and their frequency. He also proposed multiple indexing (both lexical and social) of individual memories. This multiple indexing suggests avenues for dealing with interactions of social and phonological percepts, as argued at length in Johnson (1997). With their fine-grained representations and overt modeling of social factors as well as phonetic factors, exemplar models have many

attractions for researchers in sociophonetics. An impressively thorough review of sociolinguistic results pointing towards exemplar theory is provided in this issue by the Foulkes and Docherty paper.

The purpose of this commentary is to take stock of exemplar models, both their advantages and their weak points. In doing so, it will be useful to begin by a comparison to the previous standard technique for sociophonetic analysis, Varbrul. With rigor and precision, Varbrul brings together a set of working theoretical assumptions about variation. It was a major breakthrough when it was developed, and any subsequent model is accountable for its successes.

2. Varbrul and exemplar theory as theoretical approaches to variation

According to exemplar models, variation is explicitly represented in the cognitive system by distributions of remembered examples. These distributions are built up gradually as example after example is experienced and remembered, and each example can be multiply indexed by context or function. This approach to the representation of variation is a radical departure from that taken in Varbrul, which has dominated quantitative research in sociolinguistics—including not just sociophonetics, but other areas as well—for the last two decades.

The Varbrul program implements a model that derives from the formalism of Chomsky (1965) and Chomsky and Halle (1968). These used transformational rules organized in a modular feed-forward model to describe the relationship of underlying representations to outcomes. The model is modular because each level of representation has its own descriptive apparatus, and communication across levels is highly restricted. It is feed-forward because information from the lexicon is passed to the phonology, and information from the phonology is passed to the phonetics, without any information being sent back up from any more peripheral level to any more abstract level. Specifically, Chomsky and Halle (1968) proposed that an ordered set of rewrite rules accepts abstract lexical representations as input and creates fully specified outputs, which include noncontrastive allophones and *n*-ary feature values. The fully specified outputs take the form of strings over a universal discrete phonetic alphabet; all gradient variation is attributed to universals of human anatomy and physiology.

Coverage of social effects on language outcomes was conspicuously lacking in the original formulation. However, sociophonetics shares with generative grammar a concern with the relation of lexical representations to pronunciation patterns. To capture social effects on pronunciation variants, Varbrul generalized the rewrite rules to cases of socially conditioned variability. It does so by placing probabilities on the rules, where the probabilities are conditioned by external social factors. Some key papers about Varbrul are Sankoff and Labov (1979) and Sankoff (1988); see also Paolillo (2002) for a textbook overview.

Thanks to the rigorous formulation of Varbrul (as a generalized linear model), it supports both generation and analysis. That is, it can output probability distributions of predicted results, or it can estimate underlying parameters from observed data sets. In fact, Varbrul supports analysis better than non-probabilistic generative models, as the use of probabilities permits optimization of an analysis with respect to a data set. I emphasize this point, as it is fundamental to the kind of research strategy that Varbrul can support. By using Varbrul, it is possible for researchers to make quantitative predictions, which drive the next round of data collection and analysis. Future proposals concerning the technical toolkit of sociolinguistics must support the same research strategy. This requirement is all the more urgent as researchers venture into models using network structures, as these can have very surprising and counter-intuitive behaviors.

Varbrul was never to my knowledge billed as a cognitive model. However, results on individual variation obtained using Varbrul have cognitive implications, since the regular patterns of linguistic behavior that they describe are generated by the cognitive systems of the speakers. As research on the cognitive implications of sociophonetic variation has gotten underway, the statistical assumptions of Varbrul have proved to limit its applicability.

Here are some of the limitations. First, Varbrul uses categorical variables, predicting one dependent variable from one or more independent variables. For example, Labov (1972) showed that the rate of /r/ dropping (with /r/ treated as categorically present or deleted) is a function of gender, class, and speech register. In a technical formulation, these are categorical variables which control the probability associated with the rule /r/-> 0.

As a consequence, Varbrul is not well suited for handling gradient data, a term which I will reserve for continuously variable data (in contrast to Foulkes and Docherty, who also describe variable categorical data as "gradient"). Canonical examples of gradient data include not only subsegmental variation, as reviewed in Foulkes and Docherty's Section 2.2.3, but also some suprasegmental effects, such as the pitch range variation studied in Liberman and Pierrehumbert (1984). To fit gradient, or continuously variable, data using Varbrul, it is necessary to coerce the data into a categorical format. Of course it is always possible to discretize data; even the speech recordings we all work with are digitized. However, when continuous data is categorized, information is lost which can be critical for understanding the underlying mechanisms of the system. One of the central issues in cognitive research on phonology and phonetics is categorization and gradience. Whether particular patterns of variation are categorical or gradient is hotly disputed. Understanding the typology and nature of gradient, language-specific, phonetic patterns has been one of the major advances in phonology and phonetics during the past two decades. Recent work takes up the question of how phonological categories are bootstrapped from phonetic experience (Maye & Gerken, 2000; Maye, Werker, & Gerken, 2002), and how language-specific alignments between phonological categories and phonetic cue structures can arise (Ham, 2001; Kochetov, 2002; Pierrehumbert, 2004). All of these advances depend on treating continuous variables as continuous.

Representing gradience as such is also desirable because the power of continuous mathematics can be exploited to extract more insights from fewer data points. Metric spaces provide a way of relating varied data to each other, so that it is not necessary to collect a separate statistical sample for each distinct outcome. A case in point is provided by Hay's et al.'s use of the Pillai score, a metric on the distinguishability of word pairs. This metric allows a large number of different word pairs to be treated together under the rubric of a single parameter.

A second limitation of Varbrul is that, like all other regression models, it presupposes that there are no significant interactions amongst factors. As noted in Labov (2001), social factors are very often correlated with each other. In the social sciences, a variety of methods for dealing with correlated factors have been developed. Either factors are fit incrementally until the statistical resolution of the data set is exhausted (as in stepwise multiple regression), or factors are bundled together (as in principal components analysis). So this limitation is not really fundamental. A more serious limitation, discussed in Mendoza-Denton, Hay, and Jannedy (2003), is interactions between internal and external factors. The example they take up is an interaction between lexical frequency and ethnicity of referee in the speech patterns of Oprah Winfrey. This interaction plainly places their data set outside of the range of validity of the Varbrul model. So we may ask ourselves whether this example is a bizarre artefact of a deliberately staged performance, or whether we believe it to be common and natural, a question to which I return below.

Exemplar models do not share these assumptions. As applied to speech, the lowest level of description is a parametric phonetic map rather than a set of discrete categories. Density distributions on this map are built up incrementally from effective exposure (where effective exposure is a function of actual exposure as well as cognitive factors such as attention and memory). For these reasons, an exemplar approach is obviously suited to model learning of extremely detailed language-specific or dialect specific patterns, such as the differences between Newcastle and Derby dialects presented by Foulkes and Docherty. It is equally suited to modeling gradual changes, such as those presented by Harrington. The fact that the Queen's pronunciation of the final vowel in "happy" changed gradually over the decades of her adult life can be captured using incremental updating of the density distribution for the vowel which is imputed to her cognitive system. The fact that the vowel changed in height more than in frontness—whereas the population average changed in frontness as well—raises incisive questions about whom she paid attention to, and in what respects. These questions would not even arise without analyzing the quantitative interaction between two parametric dimensions.

Exemplar models do not require an a priori decision about which variables are independent and which are dependent. Phonological categories, as represented in the mind, are viewed as clusters of similar experiences. The same powerful and fine-grained representational apparatus is provided for social categories. This means that the approach is very suited to the analysis of the relations between phonological patterns and constructed social categories. Though age and gender might be viewed as prior, or given, groups such as "people from Corby", "Nortena gang members" and "geeks" clearly represent social clusterings induced by patterns of social interaction. Some of the most interesting results in sociolinguistics relate exactly to such social

structures; Foulkes and Docherty discuss Dyer's (2002) results on the reanalysis of Scottish pronunciation patterns as Corby pronunciation patterns. Mendoza-Denton (1996) reports a relation of eye-makeup style to pronunciation in Latina gangs in Palo Alto. The back /u/ variant used by female geeks is documented in Habick (1991). By including social categories and phonological categories in a unified mental architecture, it is possible to model interactions between them in speech perception, as discussed in Johnson (this issue).

Results such as these put results on age and gender in a new light. As Johnson and Foulkes and Docherty show, correlations between speech patterns and gender cannot be reduced to inevitable biological factors. Although the affordances of the biological system shape the way that gender is manifested in speech, the interaction of social convention with biological factors could, in principle, take many different forms. Both authors review numerous studies showing that biological tendencies, such as relatively high F0 for women, can become exagerated and socially entrenched; or in other words, that the socially constructed categories can be parasitic on the biological tendencies. Although neither paper discusses this possibility, compensatory social norms are equally possible. For example, women aspiring to positions of authority could well lower the larynx, an effortful gesture which would lower both the formants and the f0, and make them sound more like men. Third, there is the possibility that speech traits that index gender can arise essentially arbitrarily, like any other type of dialectal difference. Foulkes and Docherty's Fig. 8 provides an example of this kind; clearly, preaspiration of word-final /t/ functions as a marker of gender in Newcastle but not in nearby Derby. Another case of a similar character is discussed in Pierrehumbert, Bent, Munson, Bradlow, and Bailey (2004), Vowel patterns of GLB (gay/lesbian/bisexual) speakers are shown to differ, on the average, from those of straight speakers in a way which must be learned and has a certain arbitrariness. A very striking result on gender as a social construct is also discussed in the contribution by Foulkes and Docherty. Primary caregivers use different phonetic variants in addressing young male and female children. Although some people might imagine that such behavior is a deliberate and calculated response to social norms, almost an effort to teach the children to take on their proper social roles, it seems more likely to me that it is unconscious and automatic. It is reminiscent of the automatic use of "motherese" in speaking to small children. It suggests that all gender roles are embedded in social relations, and that the exact nature and strength of social bonds inherently modulates speech production.

The same issue shows up again in a different guise in the Harrington paper, which features a careful effort to factor the effects of historical change from the intrinsic effects of aging. The sociolinguistics literature presupposes that there are two intrinsic (or biological) effects of age, and teasing them apart is already a challenge. One is the aging of the vocal tract apparatus itself. Changes in the vocal folds or vocal tract length or shape would have audible consequences, and insofar as these follow regular patterns, they should contribute to the percept of the speaker's age. The other is neural plasticity. Prior to work such as Harrington's study of the Queen, it was widely assumed that phonetic patterns were laid down in youth, and that adults were incapable of changing them because they were past the critical period for language acquisition. Because of this (putative) biological effect, the speech patterns of older people were viewed by sociolinguists as a window into the past history of the language. The speech patterns could also, one imagines, be used by ordinary people as markers of social identity, in the same way that an outmoded style of dress might serve as an marker of which generation someone belongs to.

Harrington and others have now demonstrated the possibility of phonetic learning into adulthood (though such learning is probably less fast and reliable than learning in childhood). A corollary is that intrinsic effects of aging no longer determine the whole phonetic picture. People could compensate for changes in the vocal tract apparatus. They can, as he shows, update their dialect much as they update their dress style by adopting some but not all the characteristics of new fashions. But on other other hand, it would also be possible for people to exagerate the effects of age, much as they exagerate the effects of gender. From the world of gender, we learn that people are sometimes motivated to exagerate their authority or to exagerate their vulnerability, and the social construct of aging might display the same kind of variation. Lastly, the possibility of older people forming peer groups that adopt a distinctive style of speech is not to be excluded either. For example, the marketing strategy of the Harley-Davidson corporation is to promote biking as medium of social networking for baby-boomers. With their family responsibilities lessening in their 50's and 60's, social identification with a biking club could, we surmise, give rise to social markers in the speech of Harley-Davidson baby-boomers.

Pressing this line of reasoning even further brings us back to the issue of what kinds of interactions amongst factors are natural and to be expected in sociolinguistic research. A staple of cybernetic thinking is the idea of circular causality. Circular causality arises in systems with feedback loops; A influences B, and B also influences A. In such systems, neither A nor B is the root cause of the patterns observed. Instead, the stable states of the whole system are the result of the exact interaction of the influences. Well worked out examples of circular causality include the neural circuits used to implement logical operators, entrainment of rhythmic actions, and the dynamics of predator-prey populations. Language groups provide a prima facie example of circular causality, in that people who speak the same language tend to group together and interact with each other, and people who group together and interact with each other tend to imitate each other; Welkowitz, Feldstein, Finklestein, and Aylesworth (1972), Giles, Mulac, Bradac, and Johnson (1987), Burgoon, Stern, and Dillman (1995), Nye and Fowler (2003), and Shockley, Sabadini, and Fowler (2004) are just a few of many studies showing automatic entrainment of phonetic properties amongst interlocutors. Given the existence of a positive feedback loop through the speech chain, small imitative effects can build up over time and give rise to distinct dialects or languages. To identify possible interactions amongst factors, we can trace loops through the social network of speakers and hearers, meticulously delineating the flow of information through the speech perception-production system as we go. In this light, interactions between lexical and social variables are not at all surprising. In fact, a surprising and fascinating array of such interactions is presented in the contribution by Hay et al.

3. Why have phonology?

Given the myriad attractions of exemplar theory, it is a good idea to recall why linguistic theory has phonology at all.

The phonological principle states that languages have basic building blocks, which are not meaningful in themselves, but which combine in different ways to make meaningful forms. Rearranging the same blocks results in a qualitatively distinct form with the potential for a completely different meaning (as in rearranging /taps/ to yield /spat/. The number of basic building blocks is a great deal smaller, some three orders of magnitude smaller, than the number of distinct words. The dominant patterns in speech processing (both speech production and speech perception) apply across the board to all words sharing specified structural properties. For example, if I as an American English speaker create a new word /sklati/, I immediately know that the /t/ will usually be flapped in fluent running speech.

The phonological code provides a temporal organization that facilitates the extraction of words from the speech stream. By nine months old, babies have already learned to use statistics over phoneme transitions and language specific prosodic templates to decompose the speech stream (Jusczyk, 2000). Without such bottom-up cues for word boundaries, it is most unclear how access to the lexicon, known to occur through incremental competitive activation of lexical items, could unfold at the speed and accuracy that are observed. The existence of a phonological coding level in between the speech signal and the lexicon is also evidenced by results on phonotactics and lexical neighborhood densities in speech perception. On the average, words with many lexical neighbors (that is, words which are minimally different from many other real words) have high-probability phonotactics. However, the correlation is not perfect, and Vitevich and Luce (1998) and Vitevich, Luce, Pisoni, and Auer (1999) and designed an experiment in which they are varied independently. They found that words with likely phonotactics are recognized faster, but words with many lexical neighbors are recognized more slowly. For two highly correlated factors to have effects in opposite directions in this way, it is necessary to posit a separation in the levels of representation at which the effects occur. On the production side, an equally powerful argument was already made in Shattuck-Hufnagel (1979) on the basis of speech error data. The statistical patterns of transpositions errors reveal the existence of a phonological buffer, into which lexical word forms are copied for production planning.

Historical changes were previously thought to apply in two ways. So-called Neogrammarian sound changes, originating as shifts in the phonetic implementation of phonemes, apply across-the-board to all words equally. Changes can also propagate through the lexicon, as some words are categorically restructured by analogy to others. The existence of changes which are simultaneously phonetically and lexically gradual has been a point of much dispute, though some such cases have now been put forward in some detail (e.g. Phillips, 1984). This

controversy should not obscure the generalization that the lexicon is collectively affected by historical changes. Historical change does not have the character of random drifts of the pronunciation patterns for individual words. If it did, the phonological principle would not be in force some 100,000 years after the invention of language; instead, each word would be an individual point somewhere in phonetic hyperspace. As discussed in Pierrehumbert (2002), the assumption that speech processing necessarily involves phonological parsing has the consequence that all word-forms must be constructed from the same inventory of phonological building blocks, even if within-category variation related to contextual factors also exists.

Word-forms and meanings are poorly correlated with each other. As word sets such as {"yellow", "green", "bellow", and "scream"} indicate, similarity in similarity in form ("yellow" ~"bellow"; "green" ~ "scream") does not predict similarity in meaning ("yellow" ~ "green"; "bellow" ~ "scream"). Similarity in meaning does not predict similarity in form. Though the correlation of form to meaning is now known to be slightly positive (somewhat undercutting Saussure's famous doctrine of the arbitrariness of the sign) the correlation coefficients reported in two quantitative studies are very low (Rapp & Goldrick, 2000; Shillcock, Kirby, McDonald, and Brew, 2001). What is the mechanism maintaining this low correlation? Everything else being equal, any mutual reinforcement of meaning and form should have advantaged systems in which these properties were highly correlated over the course of linguistic history, much as evolution has given rise to correlations of form and function in the various body parts of animals. The conclusion is that the some functional factor must supply a countervailing force, to decorrelate them. This issue is discussed at more length in Beckman and Pierrehumbert (2004), in which phonology is discussed as a hidden layer in the neural network system which implements language in brain.

Beginning at about 17 months, the rate at which infants learn new words begins to accelerate, and older children and adults display the striking capability of "fast mapping", or the ability to learn a new word from very slight exposure. Having learned a word, they can recognize it again in a new situation or spoken by a new speaker; see Bloom (2000) for a survey of the major findings. This is the flip side of results on episodic aspects of word learning. Though people may remember words significantly better with supporting cues in situation or voice characteristics (as discussed in Goldinger, 1998) they remember words astonishingly well without any such cues. They are far better at learning new words than learning new speakers. For example, in Bradlow, Nygaard, and Pisoni's (1999) study on the interaction of talker identification with word-list recall, subjects had 9 days of training on 10 speakers to achieve the needed levels of accuracy. Stevens, Williams, Carbonell, and Woods (1968) report error rates of 6% for a eight-way forced choice task on speaker identification after 4 hours of exposure. A lengthy subsequent literature in forensic phonetics continues to yield cautionary findings about voice identification to the present day, such as Bonastre et al. (2003).

These are some of the basic behaviors of language that we seek to capture by positing phonology. As a neo-generative model, Varbrul captures these features rather successfully, every bit as well as the structuralist theories of phonology advanced in other scholarly circles. It posits a phonological coding level, in both speech production and speech perception. The building blocks provided by this coding level can be combined in different ways to make new words. The mapping of word-forms to word meanings, which is separate from the mapping of phonetic patterns to word-forms, has no reason to be correlated. Systematic shifts in the implementation of the code (through historical change or style shifting) are predicted to affect all words equally, which is roughly correct, although secondary word-specific effects have now been identified. To what extent have these basic behaviors been captured by exemplar models? To what extent can they be captured?

The simplest exemplar models, such as Goldinger's (1996, 1998) application of Hintzman's (1986) MINERVA model, or Johnson's XMOD model (Johnson, 1997; Baker, 2005), do not have an explicit phonological coding level. There is no explicit treatment of chunking/boundary detection in speech perception, or of results such as Vitevich and Luce (1998), Vitevich et al. (1999) showing a dissociation between phonotactic effects and lexical neighborhood effects. It is unclear why allophonic rules which are triggered by classic, local structural descriptions (such as flapping, reliably triggered intervocalically under falling stress), have any priority in generalizability over equally strong, but more heterogeneous, groupings of words on the basis of their similarity.

Explaining the quantitative extent of facilitative interactions, such as the interaction of speaker familiarity with word recall in word recall experiments provides another challenge to exemplar theory. The existence of these effects is a key argument for exemplar models over neo-generative models, in which social information is

treated as a source of random variation that is factored out in perception en route to the lexicon. However, making these effects available at all in the model means that they are readily available. So why don't they pile up? In a dynamic perception-production model, minute biases will pile up over time to yield major patterns and correlations, as noted just above.

More generally, the basic exemplar models do not offer a clear picture of the XOR (exclusive OR) interactions which pervade perceptual processing (Beckman & Pierrehumbert, 2004). Each time one looks at a Necker cube, either one side pops to the visual foreground or the other; similarly, when parsing an ambiguous speech signal, such as [s:pir], one hears either "spear" or "Sapir", but not both, on any given occasion. In a standard speech processing architecture, such effects are captured though competition amongst abstract units. On each individual occasion, a single competitor wins. The net effect of such competition over time is discussed in depth in Wedel (2004). Categories, which drift too close and become confusable, tend to walk apart again. This happens because random variation resulting in more distinct and contrastive pronunciations is captured by the memory representations. However, random variation towards the middle ground of the two categories does not affect the aggregate representation. For example, if excessive lenition of the first syllable in "Sapir" happens to result in the erroneous percept "spear", the representation of "Sapir" will be unaffected. The overall effect is to probabilistically favor separation (or de-correlation) of the phonological representations, which compete with each other in similar semantic contexts.

Simple exemplar models do not offer any insights into Fast Mapping. A novel word, in the Hintzman (1986) model, could be produced by using its echo as a target, where the echo is the aggregate averaged similarity to existing words. With a word being a cluster of echos (just as an individual speaker's voice is a cluster of echos), a reasonable statistical sample of the word should be necessary for it to acquire any cognitive importance. Consider, just as an example, a novel word whose precedents in the lexicon are "simple" and "kindle". Sequential blends such as "sindle" and "kimple" can be readily learned from a single example. But a novel word which superimposed the /k/ and the /s/ (as a velarized /s/, a phonetically possible phoneme with a distinctive energy band in the F2 region) would be much less readily acquired.

A recent line of research has also revealed astonishingly rapid plasticity in use of phonological categories. Maye, Aslin, and Tanenhaus (2003) found this for categorical vowel shifts. German, Carlson, and Pierrehumbert (2005) report related results for reuse of American English flap as rhotic tap in imitating Glaswegian English. Artificial language learning experiments by Peperkamp and Dupoux (in press) also reveal the ability for adult speakers to learn and generalize novel morphophonological neutralization patterns. More broadly, the point is that a simple exemplar model has difficulty placing a priority on categorical learning (which is more abstract and held to be epiphenomenal) over phonetic learning (which is less abstract and the very stuff of the model). Given that this priority does characterize the adult linguistic system, it looks like Varbrul was a good place to start.

In short, the simplest exemplar models appear to be seriously deficient in handling classic findings, in both linguistics and psycholinguistics, which led to the development of phonological theory in general, and Varbrul in particular. When we consider simultaneously the successes of Varbrul and the strong points of exemplar theory, it is clear that a hybrid model is needed. Hybrid models in which a phonological coding level intervenes between the lexicon and the parametric phonetic description are already anticipated in Goldinger (1998) and laid out more explicitly in Pierrehumbert (2002). This approach imports the concept of levels of representation from generative models. It imports from exemplar theory the claim that probability distributions are acquired in great detail through experience, that they continue to be updated in adult life, and that episodic factors can influence the way that these distributions are used in speech processing. With more than one level of representation in the model, density distributions can be defined at more than one level; they can, in principle, relate any level to any other level. Hay et al. exploit this option heavily in their analysis.

For a technical formulation of the phonological coding level, two contemporary points of departure are Hidden Markov Models and recurrent networks. Hidden Markov Models (or HMMs) provide the standard control structure for automatic speech recognition systems. They can be built up in layers. For example, a word can be trained to have a probability distribution over alternative pronunciations, and the variants can have probability distributions over phonetic parameters. (See Rabiner, 1989; Rabiner & Juang, 1993, for an overview). Similarly, recurrent networks can also have multiple hidden layers. (See Plunkett & Elman, 1997). These approaches provide a starting point because they provide a statistically trainable system of

representation for sequential encoding, thus supporting dissociations of form and meaning, XOR interactions, Fast Mapping, and many other classic hallmarks of phonology. They do not provide a full answer to the scientific characterization of the phonological coding level because they have difficulties with large-scale gradient dependencies. One example, discussed Pierrehumbert (1993), concerns the effects of post-nuclear position on the phonetic properties of words. If a speaker shifts towards a soft, breathy voice quality over an entire string of words in post-nuclear position, this will affect the measurable properties of all the phonemes in the sequence. The same problem occurs in a different guise when dealing with sociostylistic shifts. For either case, the challenge for the future is to develop models that can shift the use of the parameter space over some temporal interval, while still exploiting the lexical representations and phonological structures that are still present, though manifested in a different way.

Hybrid models also promise to provide technical resources needed for a theory of historical change. They can capture changes involving within-level dynamics (for any level). Equally, they can describe changes involving vertical dynamics (reinterpretion of superficial patterns in terms of the representational apparatus of more abstract levels). The existence of changes that are both lexically and phonetically gradual, is not surprising, given that phonetic distributions for specific words can be defined in the model. However, the behavior of the entire perception-production system is structured by the phonological coding levels.

In summary, the traditional arguments for a phonological level of representation are valid. We need phonology to explain basic findings in the structure of the lexicon, psycholinguistics, and historical change. The first and simplest exemplar models provided a valuable challenge to the field by revealing the weak points of neo-generative approaches to sociophonetic variation. However, the future lies with hybrid models, which have multiple levels of representation (like neo-generative models) while also having explicit mechanisms for statistical learning and situational indexing (like exemplar models).

4. Learning and frequency in the cognitive system

One of the original motivations for exemplar theory was its natural treatment of effects of frequency, which are pervasive in language processing. For example, high frequency words are recognized faster and under more adverse signal conditions than low frequency words. High-frequency phonotactics speeds lexical access. The frequency bias in perception means that infrequent phonemes can only survive if they are robustly distinct from high frequency competitors; otherwise, the frequency differences are magnified through use and the weaker member of the competitor tends to be lost historically.

Such effects are naturally captured in exemplar models through the very processes of encoding and memory. More frequent categories acquire a more substantial cognitive representation, simply because tokens of these categories are (by definition) encountered more often. Thus, the representation of frequency is intrinsic to the processing in the system, and no special mechanisms need to be posited. Frequent categories are advantaged in speech perception, because speech perception involves competition amongst alternative classifications of the same physical stimulus. This competition plays out in exemplar theory through the cumulative force of the exemplars in the similarity neighborhood of the stimulus; if there are more exemplars, the cumulative force obviously tends to be greater. Warren, Hay, and Thomas (forthcoming) rely on this property in their exegesis of a frequency bias in the original SQUARE-NEAR data set.

Exemplar models involving multiple levels of representation or types of indexing naturally have frequency effects everywhere. In the phonological area, they make it possible to discuss the frequencies of different parametric outcomes, of different prosodic or segmental configurations, of categorically distinct representations of words, or of different words competing in the analysis of the same signal. They also make it possible to talk about the amount of experience with different speakers or social groups. Further, the frequencies relating to the perceptual system are not necessarily the same as those relating to production. Native listeners have experience perceiving more different dialects or idiolects than they ever undertake to produce. Even when they have mastered different sociostylistic systems in production, their intention to perform the various styles can have very different statistics from the frequencies with which they encounter these styles as listeners.

A not uncommon criticism of exemplar models is that the frequency effects are excessive, with the models predicting that more frequent means more, period. However, this is not actually a generic prediction of

exemplar models. Particularly in models with multiple levels of representation, effects can play out in surprising ways. The reasons for these surprises range from obvious to subtle. Here are some of them.

4.1. Ceiling effects

It would be astonishing if human neural circuitry proved to encode and store every single experience separately as such, like a character in a Borges story. Instead, remembering experiences involves updating or strengthening neural circuits. Extremely similar experiences would impact the same circuits, and the cumulative effect of exposure need not be linear. In fact it is very unlikely to be linear. Saturation effects for the memory of extremely high-frequency types of events are only to be expected.

4.2. Salience

Clusters of exemplars do not reflect undifferentiated raw experience, but rather experience as it has been encoded and stored. The role of differential attention in coding and memory is well-documented in psychology, and attention is not a simple function of frequency. In the widely used preferential looking paradigm for infants, the infants appear to pay the most attention (as indexed by eye fixations) to events which are right at the edge of the learning envelope; neither so familiar that they are boring, nor so novel that they cannot be coded and assimilated (Jusczyk, 2000). More generally, it appears that people are adapted to events that are most informative. Events that are attended to are in turn more likely to be remembered.

The most informative events are not, in general, the most frequent ones. Everyday examples illustrating this point are easy to come by. For example, if you bike past a certain grocery store every day on your way to work, you will quite likely cease to pay attention to it, and be unable to report even a short time later what specials were advertised in the window. However, if you pass a highly novel event, such as a hot-air balloon landing in the parking lot, you might remember the lettering on the balloon for a long time. Exemplar models are not sensitive to frequencies of ambient events per se, but rather to frequencies of memories. In between physical experience and memory lies a process of attention, recognition, and coding which is not crudely reflective of frequency.

4.3. Dissociations between perception and production

The existence of double dissociations between the perception and production systems leads to the conclusion that these systems are separate but highly coupled in normal adults. Examples include the documentation of distinct perception and production vocabularies in children (reviewed in Menn & Matthei, 1992) and the findings of two kinds of near-mergers in the sociolinguistic literature. In classic near-mergers, as discussed in Labov (1994), speakers fail to perceive differences that they produce. In Warren et al. (forthcoming), they fail to produce differences that they are capable of perceiving. The existence of separate perception and production modules means that each carries its own frequency information. Further evidence from psycholinguistics and from clinical populations is presented by Shallice, McLeod, and Lewis (1985), Caramazza and Miozzo (1997) and Martin, Lesch, and Bartha (1999). The qualitative dissociations that lead us to postulate these modules can be viewed as particularly extreme, clear, cases of frequency dissociations. However, the internal structure and propagation of frequency information on the two sides could also, in principle, differ. In production as well as perception, error patterns can be used as diagnostic information about the architecture. Results on speech errors provide intriguing hints of anti-frequency effects. According to Stemberger (1991), in speech errors, /k/ is more likely to substitute for the more frequent phoneme /t/ than /t/ is for /k/. This finding could, in principle, be incorporated into the model by positing a production analogue to information-based salience in perception. That is, if rare events are disproportionately salient to the perceptual system, it also seems possible that the intention to produce a rare event could give rise to a more focused, attentive, and energetic effort. However, all details remain to be worked out.

4.4. Facilitation and competition

One of the most powerful arguments for a phonological coding level, as already discussed, is the finding by Vitevich and Luce (1998) and Vitevich et al. (1999) that high-frequency phonotactics facilitate word recognition, whereas dense lexical neighborhoods slow word recognition. Phonotactics and lexical neighborhood density are highly positively correlated. This apparently paradoxical result is naturally explained by considering the activation and competition at each level. For the relation of the speech signal to the phonological coding, a high-frequency phoneme transition competes with other phonological analyses by its robust, detailed cognitive representation. If the speech is encoded faster, then the next level of analysis (access to the lexicon) can also proceed faster. For the relation of the parsed speech signal to the lexicon, the different words in the lexicon are all competing to be recognized. The more different competitors there for the same stretch of material, the longer if will take for the competition to resolve itself.

This example provides a template for analyzing other frequency effects. If many things gang up together in a process, they facilitate it. If they compete with each other in a process, they will cause delays. This basic observation leads to many additional predictions. For example, ample experience with a dialect should and does facilitate recognition of that dialect (cf. Clopper and Pisoni, 2004). On the other hand, knowing many speakers of a dialect could make it harder to recognize one single speaker of that dialect.

4.5. Interactions of frequency with category structure

Psycholinguistic evidence indicates that phonological categories can be projected bottom-up from the speech signal through cluster analysis. If we apply the mathematics of computational linguistics or machine learning to elucidate such results, category learning becomes a problem in statistical discrimination. It is better to attribute two distinct clusters than only one to a particular region of the phonetic space if the two-cluster hypothesis provides a significantly superior account of the data than the one-cluster hypothesis. An account is significantly superior if it models the variation in the data to a significantly greater extent. (The significance threshold is a parameter of the model, just as the cutoff for a reportable significance level is a strategic decision in scientific induction.) Learning of the phonological grammar can be treated in the same way, as clustering over words in the lexicon, or over morphologically related families of words (Pierrehumbert, 2003).

The implicit role of statistical clustering in phonological learning leads to the prediction that learning should exhibit interactions amongst frequency, sample size, and conceptual distance. Consider, for example, the task of learning an additional phoneme besides /n/ (which, for the sake of argument, I assume to be known already). If the additional phoneme is frequent, if it is phonetically extremely distinct from /n/, and if the learner has an ample speech sample, then it is extremely probable that the new phoneme will be learned. If it is rare, it is similar to /n/, and the sample size is small, then it is not likely that it will be learned. These factors interact with each other, so that learning is impaired to the extent that any and all of the factors are disadvantageous.

These interactions have not been taken into account in standard phonological models, which assume that the level of abstraction in the system is uniform. For example, almost any phonologist who posits a categorically palatal variant of /k/ in "key" would (by symmetry) posit an analogous variant of /ɡ/ in "ghee". Positing a flap category, for "butter" could lead, by analogy, to a lenis /v/ category for the equally lenited /v/ in "ever". However, mathematical learning theory means that these putative phonological entities cannot be entirely taken for granted. It is possible that frequent phonemes actually figure in the cognitive system as families of subcases, distinct by their prosodic, segmental, or social context. For rare phonemes, fine-grained learning of many subcases might not be possible. Classical phonological theory sketched a ladder of abstraction, from extrinsic allophones up to archiphonemes (or natural classes of phonemes). Extrinsic allophones were the discrete elements standing closest to the phonetic surface, representing intentions to articulate a particular segment in a particular way in a particular context. Archiphonemes were abstract classes defined partly on the basis common functions of contrastiveness, not just common phonetic substance. Where the cognitive system is poised on this ladder in any particular case is not always self-evident. Critical evaluation of the representational structure is thus a prerequisite to any successful model of frequency effects.

To summarize, hybrid exemplar models by their nature have frequency effects everywhere. However, more frequent does not necessarily mean more. Results in psycholinguistics and speech processing reveal many different kinds of frequency effects. The existence of saturation effects, salience effects, and perception-production dissociations need to be taken into account. We can't take for granted that the entities posited by linguists always correspond transparently to those used in speech processing. The relation of any given experimental task to the levels of representation in the system needs to be carefully assessed, in order to correctly predict whether frequency will be manifested in facilitation or in competition.

5. Social categories

The papers in this special issue concern the relationship of phonological categories to social categories. Social categories are in many ways analogous to phonological categories. In a population-level phonological model, each linguistic agent has a cognitive system, which includes the speaker's knowledge of phonology and phonetics. Each agent also has a social position, which can be represented as a matrix of social contacts for that speaker in the population. The entries in the matrix represent the strength of the connections, which are estimated from empirical data, such as survey data or records of the frequency of interaction. By using three-dimensional rather than two-dimensional matrices, the different nature of different contacts can also be coded (see Wasserman and Faust (1994) for a useful survey of standard methods). The cognitive and social systems are tied to each other, because the phonological system indirectly reflects the cumulative linguistic experience with the social network. They are also tied to each other by the role of social affinity in language learning. Although detailed quantitative longitudinal data on language learning are not available, qualitative results appear inconsistent with the hypothesis that sheer frequency of exposure is sufficient to explain the acquisition of socially differentiated patterns. Nobody can learn patterns in the absence of relevant exposure, but with such exposure, acquisition appears to depend on social identification. Thus, preschool children cared for at home begin by learning the caregivers' dialect, but once in school, they shift rapidly to the patterns of their peer group, even if they continue having many hours of contact with their same caregiver at home. In sociolinguistic studies of high school students, differentiated patterns for subgroups develop despite many hours of common linguistic experiences in class and through the media (Habick, 1991). Nettle (1999) in fact suggests that in hunter–gatherer communities, clan-specific speech patterns had functional value, because they permitted people to recognize others who could be expected to cooperate in long-term relations. All these results suggest a picture in which people evaluate potential interlocutors socially, and then use this information to modulate both the frequency of contact and the extent of phonetic accommodation.

In Varbrul, the social categories are just as categorical as the phonological categories; the difference is that they are external (representing observations of the scientist about the population) whereas phonological categories are internal (imputed to the minds of individual speakers). Of course, the phonological system could hardly involve more canonical categories, since word minimal pairs are perceived as qualitatively distinct. A synthetic signal which is half-way in between two words is not perceived as half-way in between those two words in meaning. For example, if we manipulate the consonant of /kau/ to bring it half-way to /sau/, the result is not half-way in between a bovine and a pig in its meaning. Instead, the ambiguous signal will be perceived sometimes as "sow", sometimes as "cow". The way that the speech perception system locks onto a single logical interpretation of the speech signal on almost every occasion makes phonological processing almost the archetype of a categorical cognitive system.

Exemplar theory provides a way to include social categories as internal factors in the linguistic model. These categories could be acquired, in much the same way that segments are acquired, through cluster analysis over perceived properties of people and social interactions. Thus, it is tempting to import the traditional social variables of sociolinguistic research directly and wholesale into an exemplar-based sociophonetic model, simply by imputing them to the minds of individual speakers as internal categories. Some of the papers in this volume read this way, though I am not entirely confident what their authors had in mind. However, I feel that this move needs to be made with some caution.

If social categories do exist, in the same sense as phonological categories, this implies the existence of a cognitive map of social parameters. Social categories would then be learned by each linguistic agent through

cluster analysis of experienced social interactions. This picture is not implausible, but it raises many issues. One issue is what the underlying parametric dimensions are.

For example, the existence of two biological sexes is one of the affordances of the social cognitive system. Does this mean, however, that the male–female distinction delineates a dimension, along which all gender-atypical people have their proper place? Though many researchers have jumped to this conclusion, it is not at all obvious that it is correct. Many GLB (gay/lesbian/bisexual) speakers would argue that gay men are not effeminate men, and that lesbian women are not masculine women; in short, that they, and their communities, differ from prototypical men and women along other dimensions than the male–female one. See Pierrehumbert et al. (2004) for further discussion. A further question is whether these additional dimensions are purely socially constructed, or whether they are adapted from other intrinsic or innate social dimensions.

Second, given the rate at which speech is processed, and the length of time it takes to acquire adult phonological competence, the sample size for phonological learning is immense. The effective sample size for social category formation is orders of magnitude smaller (counting either by distinct people encountered, or even distinct episodes of interaction). Though the sample size for some social categories (such as "my mom", "men", "women" or "my gang") is clearly sufficient, sampling issues alone suggest that social categories should be fewer and/or learned worse than phonological categories. The results of Clopper and Pisoni (2004), indicating rather poor ability to identify dialects, accord with this skepticism.

Another issue is that social factors could be relevant to the phonological system without constituting social categories, because social factors could affect the propagation of phonetic characteristics through the population, without being observable by every speaker individually. For example, the Great Northern Cities dialect in the United States has moved historically city to city, without covering the intervening countryside. The reason appears to be work-related migration patterns. However, an individual city-dweller talking with his own friends could be completely unaware of either the history of his dialect or of where, exactly, it is spoken. In short, even as we provide for internalization of social variables in exemplar-based models, we still need to allow for the possibility that some variables remain external. Sociolinguistic field studies have already established, in fact, that the way social variables effect performance on speech tasks can differ greatly, depending on the extent to which speech patterns relating to those variables have been internalized and stereotyped (see the literature review in Labov, 1994).

To summarize, population-level exemplar models provide for two kinds of social categories or variables. Social categories can be internal to the cognitive system, in which case they must be learned as clusters over remembered social experiences. The nature of the parameter space for these memories, and the learnability of sociolinguistic categories, looks like a fruitful area for future research. But we should not suppose that every social factor is necessarily reified by the cognitive system. In addition to internal factors, we still need to allow for external social factors in the model; factors which we, as scientists, can observe in the social network, but which the speakers may not have noticed themselves. In relation to these factors, we expect to see inter-speaker variation reflecting the consequences of the speaker's social position for his/her experiences. And these consequences could themselves be complex or indirect: a social position affects not only who you talk with, but also how you talk with them. Social situations give rise to emotions, and emotions in turn affect attention and memory.

6. Conclusion

Generative models of phonology, including Varbrul, encapsulate important insights about the levels of representation in language sound structure and the primacy of categorical coding. Placing probabilities on phonological rules was an important breakthrough because it permitted linguists to develop explicit models of language data with realistic amounts of variation. At the same time, this treatment of statistical variation is too limited. By using advanced exemplar models, it will be possible not only to describe probability distributions, but also to capture their dynamics in cognitive systems and in populations. Exemplar models provide tools to explore how the cognitive system handles social categories. By permitting descriptions of both internal (cognitive) and external social factors, models of this nature support investigation of circular causality in language systems, with systems arising as nearly stable states of social networks. This expanded toolkit holds promise for explaining how cognitive and social factors interact to form language.

Acknowledgments

Thanks to the James S. McDonnell Foundation for supporting my research in this area. I am also grateful to Emmanuel Dupoux for stimulating discussions about the relation of perception to production; to Ben Munson for discussion about phonetic correlates of sexual orientation; and to the editors for many years of intellectual engagement with me on the issues discussed here.

References

Baker, K. (2005) Regular and irregular pseudoverb classification using XMOD. In *The paper presented at McWOP 10*, Northwestern University, Evanston, IL, October 29, 2005.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1–22.

Beckman, M. E., & Pierrehumbert, J. B. (2004). Interpreting 'phonetic interpretation' over the lexicon. In *Papers in laboratory phonology VI* (pp. 13–38). Cambridge, UK: Cambridge University Press.

Bloom, P. (2000). *How children learn the meanings of words (learning, development, and conceptual change)*. Cambridge, MA: MIT Press.

Bonastre, J.-F, Bimbot, F., Boe L.-J., Campbell, J. P., Reynolds, D. A., & Magrin-Chagnolleau, I. (2003). Person authentication by voice: A need for caution. In *Eurospeech 2003*.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206–219.

Burgoon, J., Stern, L., & Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge, UK: Cambridge University Press.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, *64*, 309–343.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, *32*, 111–140.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*, 441–465.

Denes, P. B., & Pinson, E. N. (1963). *The speech chain: The physics and biology of spoken language* (2nd ed.). New York: W.H. Freeman and Company.

Dyer, J. (2002). We all speak the same round here: Dialect leveling in a Scottish-English community. *Journal of Sociolinguistics*, *6*, 99–116.

German, J., Carlson, K., & Pierrehumbert, J. B. (2005). Allophonic reassignment in dialect adaptation. *Poster presented at the 150th meeting of the Acoustical Society of America*, Minneapolis, MN, October 2005.

Giles, H., Mulac, A., Bradac, J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond. In M. L. McLaughlin (Ed.), *Communication yearbook*, Vol. 10 (pp. 13–48). London, UK: Sage Publications.

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

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Habick, T. (1991). Burnouts versus rednecks: Effects of group membership on the phonemic system. In P. Eckert (Ed.), *New ways of analyzing sound change* (pp. 185–212). New York/San Diego: Academic Press.

Ham, W. (2001). *Phonetic and phonological aspects of geminate timing*. New York: Routledge.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego: Academic Press.

Jusczyk, P. W. (2000). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kochetov, A. (2002). *Production, perception, and emergent phonotactic patterns*. New York: Routledge.

Labov, W. (1972). The social stratification of (r) in New York City department stores. In W. Labov (Ed.), *Sociolinguistic patterns* (pp. 43–69). Philadelphia: University of Pennsylvania Press.

Labov, W. (1994). *Principles of linguistic change: Internal factors*, Vol. 1. Oxford: Blackwell.

Labov, W. (2001). *Principles of linguistic change: Social factors*, Vol. 2. Oxford: Blackwell.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English*. Berlin/New York: Mouton de Gruyter.

Liberman, M., & Pierrehumbert, J. B. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff, & R. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: MIT Press.

Locke, J. L. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.

Martin, R. C., Lesch, M. F., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*, 2–39.

Maye, J., Aslin, R., & Tanenhaus, M. (2003). In search of the Weckud Wetch: Online adaptation to speaker accent. In *Paper presented at CUNY conference on sentence processing*, Cambridge, MA, March, 2003.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual Boston University conference on language development* (pp. 522–533). Somerville, MA: Cascadilla Press.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

Mendoza-Denton, N. (1996). ''Muy Macha'': Gender and ideology in gang girls' discourse about makeup. *Ethnos: Journal of Anthropology*, *6*, 91–92.

Mendoza-Denton, N., Hay, J., & Jannedy, S. (2003). Probabilistic sociolinguistics: Beyond variable rules. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 97–138). Cambridge, MA: MIT Press.

Menn, L., & Matthei, E. (1992). The ''two-lexicon'' model of child phonology: Looking back, looking ahead. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 211–247). Parkton, MD: York Press.

Nettle, D. (1999). *Linguistic diversity*. Oxford: Oxford University Press.

Nye, P., & Fowler, C. A. (2003). Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, *31*, 63–79.

Oudeyer, P.-Y. (2006). *Self-organization in the evolution of speech: Studies in the evolution of language*. Oxford: Oxford University Press.

Paolillo, J. C. (2002). *Analyzing linguistic variation: Statistical models and methods*. Stanford, CA: CSLI Publications.

Peperkamp, S., & Dupoux, E. Learning the mapping from surface to underlying representations in an artificial language. In J. Cole, & J. Hualde (Eds.), *Laboratory phonology*, Vol. 9. Berlin/New York: Mouton de Gruyter, to appear.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, *60*, 320–342.

Pierrehumbert, J. B. (1993). Prosody, intonation, and speech technology. In M. Bates, & R. Weischedel (Eds.), *Challenges in natural language processing* (pp. 257–282). Cambridge, UK: Cambridge University Press.

Pierrehumbert, J. B. (2000). The phonetic grounding of phonology. *Bulletin de la Communication Parlee*, *5*, 7–23.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology*, Vol. VII (pp. 101–139). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimation and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: The MIT Press.

Pierrehumbert, J. B. (2004). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 115–154.

Pierrehumbert, J., Bent, T., Munson, B., Bradlow, A., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production. *Journal of Acoustical Society of America*, *116*(4), 1905–1908.

Plunkett, K., & Elman, J. L. (1997). *Exercises in rethinking innateness: A handbook for connectionist simulations*. Cambridge, MA: MIT Press.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech processing. *Proceedings of the IEEE*, *77*(2), 257–286.

Rabiner, L. R., & Juang, B. (1993). *Fundamentals of speech recognition*. Englewood Cliffs, NJ: Prentice-Hall.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499.

Sankoff, D. (1988). Variable rules. In U. Ammon, N. Dittmar, & K. J. Mattheier (Eds.), *Sociolinguistics: An international handbook of the science of language and society*, Vol. I (pp. 984–997). Berlin: Walter de Gruyter.

Sankoff, D., & Labov, W. (1979). On the uses of variable rules. *Language in Society*, *8*(2), 189–222.

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: Are speech perception and production separate processes? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *37*(A), 507–532.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for serial order mechanism in sentence production. In W. E. Cooper, & E. C. T. Walter (Eds.), *Sentence processing* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shillcock, R., Kirby, S., McDonald, S., & Brew, C. (2001). Filled pauses and their status in the mental lexicon. In *Proceedings of the 2001 conference on disfluency in spontaneous speech DiSS'01* (pp. 53–56). ISCA Archive.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422–429.

Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, *30*, 161–185.

Stevens, K. N., Williams, C. E., Carbonell, J. R., & Woods, B. (1968). Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. *Journal of the Acoustical Society of America*, *44*(6), 1596–1607.

Vitevich, M., & Luce, P. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325–329.

Vitevich, M., Luce, P., Pisoni, D., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, *68*, 306–311.

Warren, P., Hay, J., & Thomas, B. The loci of sound change effects in perception and production. In J. Cole, & J. Hualde (Eds.), *Laboratory phonology*, Vol. 9. Berlin/New York: Mouton de Gruyter, forthcoming.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Welkowitz, J., Feldstein, S., Finklestein, M., & Aylesworth, L. (1972). Changes in vocal intensity as a function of interspeaker influence. *Perception and Motor Skills*, *35*, 715–718.

Wedel, A. (2004). Category competition drives contrast maintenance within an exemplar-based production/perception loop. In *Proceedings of the seventh meeting of the ACL special interest group in computational phonology* (pp. 1–10). Association for Computational Linguistics, Barcelona, Spain, July 2004.

Wiener, N. (1948). *Cybernetics*. New York: The Technology Press, Wiley.