3

Decisions and Mechanisms in Exemplar-based Phonology

Keith Johnson

3.1 Introduction

In the leaf in Figure 3.1 you can see a branching structure—an almost crystalline organization that could be described with a clean mathematical generative formalism. Now, if we raise our gaze only a little we see a forest—diversity formed from interlaced systems of water and light, plant and insect.

We can approach language, too, from these two perspectives. Looking at the geometric regularities in the structure of the leaf is analogous to adopting a structuralist linguistic framework inspired by mathematical/physical theories of mathematics and physics. Generative phonology (Chomsky and Halle 1968) is the most prominent instance of this approach to language in the domain of phonology, adopting such familiar research strategies as idealization of the speaker/hearer and the use of formal symbolic representation of generalizations observed in linguistic data.[1]

Seeking to understand the leaf's structure in the interacting systems of a forest is analogous to approaching language from an ecological or systemic framework inspired by theories in biology and history. Phoneticians have approached language in this way for many years (Lindblom *et al.* 1984; Blevins 2004; see also Baudouin de Courtenay 1972a) ‘‘explaining'' language sound patterns in terms of phonetic tendencies in speaking and listening that operate in the history of language (Hume and

This paper benefited from comments and discussions at the 3rd Seoul International Conference on Phonology, where it was presented in June, 2005. Parts of this research were supported by NIH Grant #R01 DC04330.

[1] Chomsky and Halle (1968) used the term ‘‘grammar'' to refer both to ‘‘the system of rules represented in the mind of the speaker/hearer'' and to ‘‘the theory that the linguist constructs'' while analyzing forms produced by speakers. Although many of the patterns discovered by linguists are no doubt psychologically real (Halle *et al.* 1981) and thus are part of the cognitive ecology of language, it is nonetheless accurate to identify Chomsky and Halle's linguistic research strategy as a continuation of American structuralism—an insular approach to linguistics that is fundamentally incapable of accommodating a broader, contextualized, view of language.

Figure 3.1. A leaf

Johnson 2001*a*). John Ohala, in particular, emphasized the historical part of the equation (1974, 1981*b*).

It is important to keep in mind that the structural and ecological perspectives feed each other. For example, many of the observations about English sound patterns given by Chomsky and Halle (1968) were derived from the history of English, so that much of the explanatory power of their analysis of English was gained because the authors explicitly referred to the historical development of rules like vowel shift and velar softening. Ohala objected to their method of packing the history of the language into the head of the speaker—incorporating this explanatory power into a synchronic grammar—but at the same time Ohala was pursuing a very similar project, finding the explanation of the current linguistic state (described in formally expressed generalizations) in the historical path that the language took through an ecology made up of the physiological and perceptual contexts of speech transmission. My point here is simply that the strictest formal description of language sound patterns benefits from historical explanation, and the most ardent biological description benefits from formalized generalizations.

The topic of this chapter is exemplar-based theories of phonological knowledge. From my point of view, exemplar-based theories may be used to increase our understanding of language from an ecological perspective by providing a framework within which we can account for generalizations in language sound systems while incorporating phenomena such as historical drift and contextual variation in phonetic detail. Whereas in phonetic research we have been primarily interested in effects on language sound systems that arise from speech perception and production, the exemplar-based approach is concerned more particularly with the cognitive grounding of phonological knowledge.

The sections to follow will outline what exemplar-based phonology is (Section 3.2), discuss two decisions that must be made in exemplar-based phonology (Section 3.3), and, finally, discuss two mechanisms used in constructing an exemplar-based phonology (Section 3.4).

3.2 WHAT IS EXEMPLAR-BASED PHONOLOGY?

3.2.1 Background

The exemplar-based orientation to sensory memory has a long history in cognitive psychology, and from the work in general cognition we get some ideas about how exemplar-based phonology might look.

Richard Semon (1923) in his *Mnemic Psychology* distinguished *sensation* and *image*. For him the term ''image'' refers to memories of sensory experience that persist in neural structure and of these he says, ''Every moment of individual life adds something to the already existing sum of simultaneous engram-complexes'' (p. 171), which is to say that each instant of life adds exemplars to memory. In Semon's view then, these images of sensory experience are used in recognition, which he describes as ''partial return of the inner energetic situation which was present at the formation of the engram-complex'' (p. 180). Two points in Semon's approach are important. First, his view that each moment of life adds to the sum of images illustrates that exemplars on this view are tokens of experience not types. That is, exemplar-based models envision the storage of instances as they occur, without any abstraction at all. Second, Semon envisions that new experiences are recognized as being similar to old experiences by a partial re-experiencing of images/instances in memory. This is an early conception of an activation model of perception.

These characteristics of exemplar-based models of memory have been carried forward into modern cognitive psychology by a number of researchers. For example, Hintzman's (1986) ''multiple-trace'' memory model MINERVA (which has been applied to speech perception by Goldinger 1992) implemented a simple version of Semon's images and developed an explicit account of ''re-experiencing'' during recognition that we will come back to in a subsequent section. Here, the main point that I want to make is that exemplar-based memory models are current in cognitive psychology research and that they are considered one of the mainstream approaches to modeling memory (Baddeley 1997; Tulving and Craik 2000; Neath and Surprenant 2003).

This point is further illustrated by a series of influential papers by Nosofsky (1986, 1988, 1991; Cohen *et al*. 2001). In his models of recognition and categorization processes Nosofsky assumes that people store in memory each instance of the members of a perceptual category and that various effects in categorization performance (frequency effects, ''prototype'' effects, recency effects) all emerge from this memory storage system. This work has been very influential, and provocative.

In the study of language as well, the idea of representing linguistic categories in terms of experienced instances of linguistic objects has been a focus of study for some years now. For example, Skousen (1989) used an exemplar-based memory system to account for processes of analogy in phonology and historical linguistics. Goldinger (1992) found evidence for exemplar-based storage of auditory words in listeners' word recognition performance. Jusczyk (1993) and Morgan *et al*. (2001 and Anderson *et al*. 2003) have proposed exemplar-based models of child phonology acquisition. Johnson (1997*b*) proposed that an exemplar-based model can account for the process of talker normalization in speech perception. Coleman (2002) proposes exemplar-based phonetic representations. Pierrehumbert (2001*a*, 2003*a*) models pronunciation variation and phonological learning using an exemplar-based phonetic storage system. Bybee (1985) modeled paradigm leveling and other patterns of change in historical morphology using prototype theory, and has more recently (2001) explored instance-based models to account for both prototype effects and the role of frequency of occurrence in these processes. All of these exemplar-based approaches to phonology assume that language sound systems are represented in the set of phonetically detailed exemplars of speech that the speaker/hearer has experienced, and that phonological generalizations—the stuff of phonological rules—emerge from the detailed exemplars. The models thus implicitly entail the radical claims that phonology is represented in phonetic detail rather than in featural abstraction, and that the phonetic definition of phonological contrast is language specific.

It should be noted, however, that there is no one "exemplar theory". Skousen has an explicit model, Goldinger uses MINERVA, Jusczyk called his model WRAPSA, Morgan's model is called DRIBBLER, and I called my simulations XMOD (which may or may not be the best model, but it is definitely the best name). In addition, there are relatively few studies that test the basic assumption of exemplar-based modeling—that people remember exemplars of speech (but see Lightfoot 1989; Schacter and Church 1992; Palmeri *et al*. 1993; Goldinger 1992, 1996, 1997; Nygaard and Pisoni 1998; Goldinger and Azuma 2004). My take on this last point is that the memory literature pretty convincingly demonstrates that an exemplar-based memory does exist for sensory experience, and there is a tendency on the part of people who have read that literature to accept the assumption and start modeling. However, additional work on this topic would be welcome.

The main point of this section is that exemplar-based models of human memory have been considered for at least 100 years. Exemplar theory is not an invention of linguists—there is a large body of work out there for us to draw on and benefit from—and even among linguists a variety of exemplar-based approaches are being tested.

3.2.2 An exemplar-based approach to phonology

So we have these two general approaches to language sound systems. One, the approach used in constructing grammars and dictionaries, is to find generalizations among pronunciations and use phonetic details noted during close inspection of people's pronunciations to formulate rules that describe generalizations among pronunciations. The second, more tentative, exemplar-based approach situates language in a cognitive model of human memory by assuming that people use an exemplar-based memory system to store phonetic details. Generalizations then are computed by the talker flexibly on-demand over this large store of phonetic exemplars.

To illustrate the exemplar-based approach and how it relates to our more familiar grammar and dictionary approach we will consider a couple of analogies. First, consider the field linguist's note cards and their use in formulating phonological observations. At the outset of the linguist's exposure to a language he or she writes words, noting as many phonetic details as possible. Eventually, though, after some time spent hearing and speaking words, and some analysis of the distinctive sounds of the language, we begin to use a more abstract alphabet—a phonemic representation. As it turns out, though, sometimes this move from detailed to abstract representation is taken too early and the linguist must go back and re-elicit forms because a missing or neglected phonetic detail turns out to be important. Now, in the exemplar-based approach, the linguist's note cards are exemplars. One interesting research question is: do learners form abstract representations such that they must essentially re-elicit forms when a generalization proves to be wrong, or do these data exist for learners in detailed phonetic exemplars available for wholesale reanalysis?

Second, consider the difference between definition by extension and definition by rule. For example, among the important numbers in my life I would include my birthday, telephone number, address, and various identification and code numbers. This set of numbers must be defined by extension—simply listed on scraps of paper in my wallet or memorized. This is a very different mode of generation than the sets of numbers that can be generated by rules such as the set of all numbers less than eleven, or the Padovan sequence $[P(n) = P(n-2) + P(n-3)$ with $P(0) = P(1) = P(2) = 1]$ 1, 1, 2, 2, 3, 4, 5, 7, 9, 12, . . . In the generative view, pronunciation variation is defined by rules of assimilation, deletion, and the like, so we assume that the speaker/hearer does not need to memorize every variant encountered or produced, but instead is able to derive variants by rule the way we can derive 4410 as the thirty-first number in the Padovan sequence.

The exemplar-based approach views pronunciation more like a problem in definition by extension. We note that many aspects of pronunciation variation don't seem to fit the rule-governed approach. For example, individual talker-specific variation and dialect variation may not be rule governed. Stollenwerk (1986) illustrates this with a report of idiolectal variation in the American English [ɑ]/[ɔ] contrast. For this speaker, high-frequency words that should have [ɔ] do, while low-frequency /ɔ/-words are pronounced with [ɑ]. Stollenwerk's explanation for this idiosyncratic distribution is that the speaker picked up high frequency words in a speech community where the distinction is maintained early in life and then moved to a community where the distinction is not made and acquired the low-frequency words with [ɑ].

Obviously the speaker memorized particular variants according to the norms of her speech community at the time of acquisition, but also her interlocutors have to also be able to tolerate such variation within the individual. That is, a listener who has a dialect-mapping rule so that if speakers are from the upper Midwest we expect the [ɑ]–[ɔ] contrast to be maintained, will be confused by them. The outstanding feature of an account of variation like Stollenwerk's is that the personal history of particular words has explanatory value.

We see this also in studies of word-specific pronunciation variation. Lavoie (2002) found that *four* and *for* have different reduction patterns in connected speech (see also Pierrehumbert 2002). I explored this further by taking counts from the *Variation in Conversation* corpus of conversational speech (Pitt *et al*. 2005) as shown in Table 3.1. This table illustrates that in normal conversational speech homophones do not have identical variant frequencies, and sometimes they do not even have the same leading pronunciation. This is true whether we compare content word and function word, as Lavoie did with *four* and *for*; two function words, as in the pairs *your–you're* and *there–they're*; or two content words, as in *one–won*, *right–write*, *hole–whole*. Here we see that words may have their own phonetic history (W. Wang 1969) so that it would not be a surprise at all to find that in some future version of American English, *right* is pronounced with a phonemic glottal stop while *write* retains the older final /t/. It may be that we can describe pronunciation variation in terms of simple rules, like the one that palatalizes and devoices the /j/ of *your* and *you're* after [t̞, s̞, ʃ, tʃ], so that we can argue that the observed variation does not emerge from word-specific variation, but only that the most typical contexts of the homophones differ and therefore their most typical contextual variants do too. In an exemplar-based view, though, the frequency distribution of variants is part of the representation of the word; thus the representation needs to change very little to support a sound change from /rɑɪt/ to /rɑɪʔ/ because the prevalence of [rɑɪʔ] variants is a part of the representation of *right*.

Coming back to the main point of this section, talker-specific patterns of pronunciation as studied by Stollenwerk (1986) and word-specific patterns of variation as studied by Lavoie (2002) illustrate the type of phenomenon that exemplar-based models handle by defining language sound patterns by extension rather than by rule.

3.2.3 Recognition memory and declarative memory

Before turning to the decisions and mechanisms of the title of this chapter, I would like to mention briefly a relevant distinction that has been made in research on memory. This distinction is justified in part by the tragic case of an epilepsy patient known as Patient HM (see Gluck and Meyers 2000 for a lucid discussion). HM underwent surgery to remove the portion of his brain that was the initiating focal point for his severe epileptic seizures. The part of his brain that was removed was the hippocampus. This was done in the days before the crucial role of the hippocampus in

TABLE 3.1. Relative frequency of occurrence of the leading variants of homophones in a 100,000-word phonetically transcribed corpus of conversational English. The number of tokens of words that occur less than 10 times in the corpus is indicated

<table>
<tr><td>four
[fɔr] 53%
[foʊr] 34%</td><td>for
[fɚ] 58%
[fɔr] 16%
[fə] 8%</td><td></td><td>won
[wən] 94%
[wə̃n] 6%</td><td>one
[wən] 73%
[wəɾ̃] 12%
[wə̃] 5%</td></tr>
<tr><td>two
[tu] 84%
[tə] 11%</td><td>to
[tə] 21%
[tɪ] 11%
[tu] 10%
[ə] 10%</td><td></td><td>past (n=5)
[pæs] 80%
[pæst] 20%</td><td>passed
[pæst] 80%
[pæs] 17%
[pæʃ] 6%</td></tr>
<tr><td>eye (n=4)
[aɪ] 100%</td><td>|
[aɪ] 58%
[ə] 16%
[ɑ] 16%</td><td></td><td>write
[raɪt] 39%
[raɪɾ] 39%
[raɪʔ] 10%</td><td>right
[raɪʔ] 38%
[raɪt] 28%
[raɪ] 14%
[raɪɾ] 7%</td></tr>
<tr><td>buy
[baɪ] 96%
[bæ] 4%</td><td>by
[baɪ] 86%
[bɑ] 5%</td><td>bye (n=5)
[baɪ] 80%
[bɑ] 20%</td><td>they're
[ðɛr] 53%
[ðɚ] 15%
[ɛr] 6%
[nɛr] 5%</td><td>there
[ðɛr] 54%
[ɛr] 12%
[nɛr] 8%
[ðɛ] 5%</td></tr>
<tr><td>knew
[nu] 96%
[nɪ] 4%</td><td>new
[nu] 77%
[nə] 11%
[ni] 6%</td><td></td><td>threw (n=6)
[θru] 67%
[θu] 33%</td><td>through
[θru] 55%
[θu] 12%
[θrə] 5%</td></tr>
<tr><td>hire (n=5)
[haɪɚ] 60%
[hɑr] 40%</td><td>higher
[haɪɚ] 92%
[hɑr] 8%</td><td></td><td>wear
[wɛr] 69%
[wər] 13%
[wɚ] 13%</td><td>where
[wɚ] 36%
[wɛr] 28%
[wər] 7%
[ʍɚ] 6%</td></tr>
<tr><td>hear
[hir] 62%
[hɪr] 31%</td><td>here
[hɪr] 49%
[hir] 36%</td><td></td><td></td><td></td></tr>
<tr><td>hole (n=7)
[hoʊl] 71%
[hoʊ] 14%
[hoʊl̩] 14%</td><td>whole
[hoʊl] 81%
[hoʊ] 16%</td><td></td><td>wore (n=5)
[wɔr] 100%</td><td>war
[wɔr] 67%
[wər] 22%
[woʊr] 11%</td></tr>
<tr><td>hour
[aʊɚ] 50%
[aʊr] 29%</td><td>our
[ɑr] 72%
[ər] 13%</td><td></td><td>your
[jɚ] 70%
[ʤɚ] 6%
[jɪ] 3%</td><td>you're
[jɚ] 72%
[jɪ] 7%
[ʃɚ] 4%</td></tr>
<tr><td>know
[noʊ] 80%
[nə] 9%</td><td>no
[noʊ] 81%
[nə] 13%</td><td></td><td></td><td></td></tr>
</table>

memory was known. As a result of the surgery, HM was unable to store any new memories. His doctor had to reintroduce himself each time he entered HM's room because HM did not remember him from one visit to the next. Readers who have seen the movie *Memento* will recognize the memory deficit.

However, more careful testing of HM's memory revealed that he did have the ability to remember *some* things from day to day. In particular, researchers taught him a game that he had never played before and found that even though they had to explain the rules of the game again each day that they visited (after introducing themselves) HM got better and better at the game. He was learning how to play it, using a memory system not available in his conscious life.

This type of priming without conscious memory, and research with normal subjects that reveals a difference between implicit and explicit memory, have led researchers to distinguish recognition memory from declarative memory. Declarative memory, on this view, is made up of one's knowledge of expressible facts, the sort of knowledge you could gain by reading books. Contrasting with this is the type of memory that HM seemed to be acquiring (but without the ability to transfer it to a declarative memory store) as he played the game. This implicit recognition memory is comprised of knowledge acquired through direct experience of an event or object. It is detailed in nature but often hard to describe in words. For instance, you can recognize close friends quickly and often from very limited sensory stimulus (in a quick glance, for example), but if you are asked to describe them to someone who will pick them up at the airport you may find yourself struggling for words. Recognition memory is thus the type of knowledge you get from direct experience while declarative memory is a kind of encoded representation of knowledge that can be passed from person to person in language.

It seems inevitable that the richness and directness of recognition memory is the language-user's knowledge that underlies linguistic performance, while our description of this knowledge in grammars and dictionaries is an impoverished representation in the same way that my description of my friend to the person who will pick him up at the airport is an impoverished representation of my true mental representation of him. While grammars and dictionaries are indeed powerful representations of language, exemplar-based modeling of phonology seeks to explore a representation of phonological knowledge that may be a little closer to the richness of language as it is experienced and stored by native speakers.

In what follows, then, I will recommend answers to two key decisions that must be made in exemplar-based phonology, and then discuss two mechanisms (one old and one new) that I think are required to model successfully language sound patterns in an exemplar-based phonology.

3.3 TWO DECISIONS

The first decision that we make in implementing an exemplar-based model of phonology is that of choosing a unit of representation. Some researchers have proposed to represent exemplars of speech sounds (Skousen 1989; Pierrehumbert 2001*a*) while others suggest that exemplars in memory are exemplars of words (Wedel 2003; Johnson 2005*b*).

I assume that exemplars are "of" experiences, that the waves of sensation that we are subject to are segmented into conscious experiences.[2] Searle (1998) argues that nonconscious brain states or events like neurotransmitter release, and my belief while sleeping that airplanes can fly, are not experiences. I accept his suggestion that experiences are the product of the conscious mind. Related to this, Edelman (1987) suggests that conscious experience is generated by the interaction between neuronal maps in the brain. So, in considering what a linguistic experience is, and, thus, the level of representation in an exemplar-based model of phonology, I would suggest that in language people generally experience words and not sounds. One line of support for this decision is in noticing how speakers and listeners talk to each other about language. Nonlinguists ask about words—word meanings and word pronunciations—without noticing or commenting on sub-word regularities. For example, in southeastern Ohio, *fish* is pronounced [fiʃ] and *push* is pronounced [puʃ] (instead of Standard American English [fɪʃ] and [pʊʃ]). Of course, this is a pretty general phenomenon—high lax vowels followed by [ʃ] in Standard American English are pronounced tense in the southern midlands dialect. But this is not how speakers seem to experience this phenomenon at least in conscious experience—we talk about words not sounds. I think this makes sense from Edelman's perspective on the neural formation of consciousness because words are where form and meaning, which are represented in different neuronal structures, interact with each other. So if, as Edelman suggests, the interaction of neuronal structures is the locus of the generation of conscious experience, it would make sense that words would be the fundamental building blocks in the conscious experience of language, while sounds are much less accessible to consciousness. In light of this perspective on the speaker's experience of language, then, exemplar-based phonology should start with word-sized exemplars.

The second decision has to do with how to represent the dimensions of exemplars. Nosofsky (1986) used multi-dimensional scaling to select perceptual dimensions, and in many ways this makes great sense because the representation is compact and based on data. The alternative is for the modeler to make assumptions about how to represent speech (as e.g. Pierrehumbert 2001*a*; Wedel 2004*a*) using perceptual dimensions that may or may not be important to listeners. Nosofsky had very simple stimuli (circles of various sizes with one radial line at various angles in the circle) and his two perceptual dimensions were very highly correlated with the two stimulus dimensions. In order to make a perceptual map of speech stimuli, on the other hand, we need *many* more dimensions than this—including auditory, visual, proprioceptive, and motor-control representations. Just considering the auditory representation of speech

[2] One reviewer asks how HM, who could not consciously remember experiences after hippocampus removal, could have conscious awareness without the ability consciously to remember those experiences. As I understand it, the hippocampus is involved in recoding experience to declarative memory and that hippocampus injury does not change experience. Thus, HM still had conscious experience after surgery and his experiences could still be treated as exemplars in recognition memory, though he lost the ability to access exemplars consciously.

we could code in terms of F1, F2, F3, f0, duration, spectral shape parameters, plus dynamic representations of each of these. The first two formant values have been shown to be important dimensions in vowel perception but multidimensional scaling spaces for larger sets of phonetic segments (Winters, p. c., Heeringa, p. c.) have not proven to be so coherent. In Johnson (1997*b*) I used formant values as the dimensions of a vowel exemplar space, but in Johnson (1997*a*) I stored exemplars as auditory spectrograms of words. This rich auditory representation is realistic because it is based on psychoacoustic data, and it also avoids making assumptions about which of the many potential acoustic features should be measured and kept in an exemplar of heard speech. I think that currently this is the best approach, not because it gives the cleanest and easiest modeling results—because this is definitely not the case—but because there is not enough data-driven evidence for a more compact representation. Detailed auditory spectrograms are problematic partly because they are so detailed and one would like to be able to reduce the information stored for each exemplar to a small number of significant parameters. However, though parameterizing, the perceptual space may make our models easier to work with, the data that could guide us to a particular parameterization do not exist. So, my decision has been to stick close to the signal in the model representation of exemplars—attempting in essence to have the same memory representation support modeling of low-level speech perception phenomena like talker normalization, while also aiming to account for higher-level language sound patterns.

The ideal model of exemplar-based phonology (and perhaps the only way to make this enterprise work) is to include visual and articulatory information in (some) exemplars. I speculated (Johnson 1997*b*) that inclusion of "seen" and "self" exemplars would provide a measure of phonetic coherence in perception that is lacking in current models. This remains an important research task.

3.4 TWO MECHANISMS

This section is a discussion of two key mechanisms used in exemplar-based phonology. These two mechanisms are methods for calculating activation of exemplars in response to input and the spread of that activation in a network of exemplars. In a model of speech perception the input is a detailed stimulus token and the pattern of activation is used to determine the category membership of the exemplar. In a model of speech production the input is a desired category output and the pattern of activation is used to determine the phonetic details of the speech to be produced. In either case, we must specify a mechanism of exemplar similarity matching, and we must specify a mechanism of activation spreading.

The first mechanism in exemplar-based phonology (similarity matching) has been used in exemplar-based memory models for many years. It is reviewed here for the sake of making an aspect of exemplar-based modeling explicit. Some authors have emphasized that exemplar-based models produce "on-line" generalizations. Hintzman (1986) was perhaps most explicit about this, even in the title of his article mentioning "abstraction". There is sometimes the misapprehension that because there are no abstract category prototypes in exemplar-based models, it must follow that exemplar-based models may not exhibit prototype effects such as generalization or abstraction. This is not the case because the aggregate response of category exemplars displays exactly these generalization characteristics. The question is not whether people behave as if they have stored abstract category prototypes, but, rather, whether this behavior arises from exemplar storage or prototype storage—because exemplar-based systems do exhibit abstraction behavior. The argument in favor of the exemplar-based generalization mechanism is that people also exhibit exemplar-tuned behavior, so if some sort of exemplar storage system is needed anyway, and if such a system can exhibit generalization behavior (via a similarity matching process) then why would one posit a parallel, totally redundant, prototype system?

The similarity matching process, then, is a key feature of exemplar-based memory systems. Several similarity matching algorithms have been proposed. I will discuss the model given by Nosofsky (1986) because it is the similarity-matching model that I have used in my own work, but many people find Hintzman's MINERVA a more intuitively approachable model.

The first step in similarity matching is to calculate the euclidian distance between the input auditory spectrogram (j) to all exemplars (i). One trick here is in temporally aligning the auditory spectrograms with each other. The strategy taken in XMOD is to slide x_i by x_j (permitting all possible alignments) and let d_{ij} be the smallest observed distance between them.

$$d_{ij} = \sqrt{\sum (x_i - x_j)^2} \quad \text{Auditory distance}$$

Now each exemplar's activation is calculated from auditory distance. The amount of activation of exemplar i caused by input token j is an exponential function of the auditory distance between the exemplar and the input token. One model parameter c scales the activation and is constant for all of the exemplars.

$$a_{ij} = e^{-cdij} \quad \text{Exemplar activation}$$

Finally, the evidence that input token j is an example of category k is then a sum of the activations of all of the exemplars of category k. In this formula w_{ki} is a weight set to 1 if exemplar i is a member of category k and 0 otherwise.

$$E_{kj} = \sum a_{ij} w_{ki} \quad \text{Category activation/evidence}$$

This simple similarity-matching algorithm produces prototype behavior from a set of detailed exemplars. As we will discuss shortly, a second mechanism may tune the output of the matching algorithm causing the "prototype" of the category to shift in response to various sorts of contextual factors.

The second mechanism that is important for exemplar-based phonology is an exemplar resonance mechanism that permits activation to spread through the set of exemplars via non-phonetic properties. In this approach, each exemplar of language has phonetic properties and nonphonetic properties and similarity between exemplars on their non-phonetic aspects changes the phonetic response of the system. Thus, for example, when you hear a token of a particular word some word properties (e.g. meaning, spelling, usage) may become active and then in a resonance loop feed that activation back to the exemplar memory so that the similarity-matching process becomes weighted toward the word. Hintzman (1986) envisioned a loop of this kind that sharpens the response of the system so that even on relatively equivocal evidence the exemplar-based system will come to a definite recognition decision. In cognition research generally, resonance is one of the key building-blocks in neural modeling. Semon (1923) described a resonance loop in the generation of ''mnemic'' excitations. Grossberg (Carpenter and Grossberg 1987) calls his influential neural modeling approach the ''Adaptive Resonance Theory'', with resonance as one of the key explanatory mechanisms. And Edelman (1987) emphasizes the power and importance of re-entrant mapping between neural subsystems in the generation of consciousness.

Coming back to Hintzman's (1986) MINERVA model as an illustration of resonance with exemplars that contain phonetic and nonphonetic information, we can extend Hintzman's ideas to phonology by modeling exemplars (Fig. 3.2) as a set of phonetic and nonphonetic properties. If the input to the system is purely phonetic (though often it is not, as we will see), the aggregate similarity response (the ''echo'') from a set of exemplars that do have specification for both phonetic and nonphonetic properties will contain an *altered* specification for the phonetic properties and a *new* specification for nonphonetic properties. For example, if one group of the nonphonetic properties of the stored exemplars codes the identity of the word, and another group codes the identity of the talker, then after a number of resonance echos the system will settle into stable categorization responses for both the talker and the linguistic item.

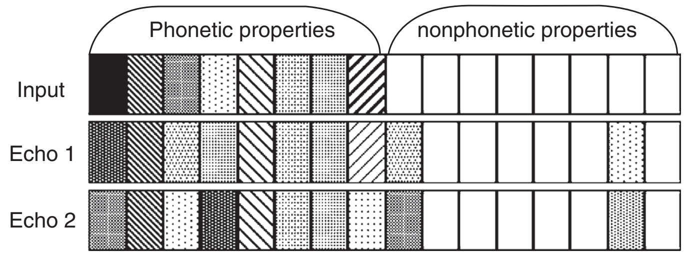

FIGURE 3.2. An input exemplar composed of phonetic property values (shaded cells) and with no nonphonetic properties specified. The first echo from a set of exemplars has a few nonphonetic properties activated and some of the phonetic property values altered. The second echo increases this trend

Given this, it is easy to see also how ''topdown'' activation could alter perception. Pickett and Pollack (1963) reported that context improves the perception of words produced in conversational speech, and Lieberman (1963) found that speakers produce clearer variants of words when the words appear in nonpredictive environments. In Hintzman's approach we suppose that some nonphonetic semantic properties are provided in the input with the phonetic properties, so that now the phonetic properties do not have to be so distinct in order for the correct word to be recognized because additional topdown information is involved in the similarity matching. In a Nosofsky-style exemplar model (Fig. 3.3) we would allow nonphonetic contextual information to alter the resting activation level of exemplars according to how well the exemplar matches the context, so that if the topic of conversation includes the notion <fish>, all exemplars that match this notion in their nonphonetic properties will get a boost of activation, and an incoming stimulus that sounds (taken out of context) more like *cot* than *cod* will nonetheless be recognized as *cod* because the overall activation of *cod* exemplars is higher in the combination of top–down and bottom–up evidence. In either Nosofsky's approach or Hintzman's we have interacting activation between phonetic and nonphonetic information that alters the perception of phonetic material allowing topdown information to change the perceptual process.

Another illustration of this comes from sociophonetics. It has been observed that listener expectations can alter perception (Johnson *et al.* 1999). One particularly striking illustration of this was shown by Strand (2000). She found that listeners were slower in naming words produced by nonstereotypical-sounding male and female voices than they were in naming words produced by speakers who sounded

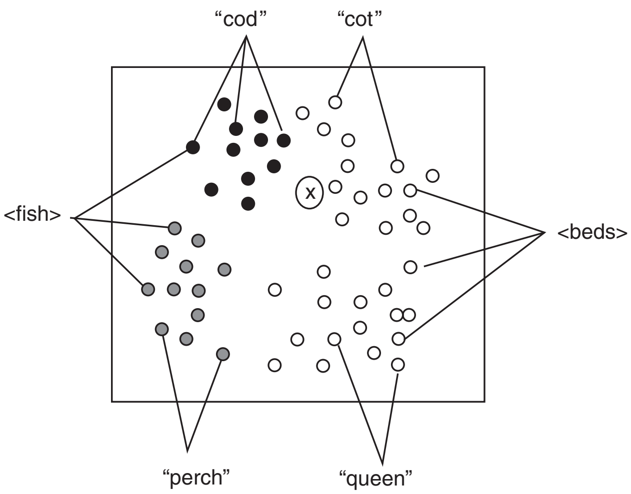

Figure 3.3. An illustration of how top–down activation might increase the activation of all exemplars related to <fish> so that a phonetically ambiguous token that sounds more like *cot* than *cod* will be recognized as *cod*

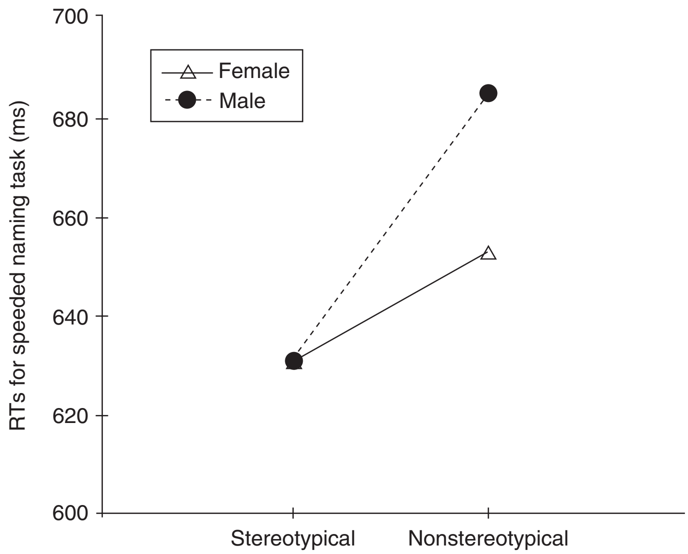

Figure 3.4. Naming times were longer for nonstereotypical-sounding voices than for stereotypical-sounding voices (after Strand 2000)

stereotypical (Fig. 3.4). The suggestion here is that voice "stereotypicality" arises in the interaction between phonetic and nonphonetic properties coded in exemplars. The resonance process is quick and decisive with stereotypical-sounding voices producing an easily classified coherent response, while nonstereotypicality results in at least momentary ambiguity in a resonance process that is thus slower to match phonetic detail to categorical representation.

As with the effect of top–down semantic information, we can envision the effect of talker information (via the acoustic signal, a visual signal, a listener bias) as producing an effect on perception by altering the resting activation levels of exemplars associated with the talker. This is illustrated in Figure 3.5. If the gender of the talker can be clearly identified, then the evidence for word identity is sharpened by reducing the amount of competition, thus supporting faster word-identification response.

Finally, the resonance mechanism in exemplar-based phonology permits phonological generalization as well. Some aspects of phonological patterning "emerge" from resonance between semantic and phonetic information. A sketch of this idea is given in Figure 3.6. The allophonic relationship between [d] and [ɾ] emerges from semantic/phonetic resonance. Forms associated with both the word "odd" and the word "odder" are activated in response to presentation of a token of "odd" because there is a semantic relationship between "odd" and "odder". At first only exemplars of "odd" will be activated because the input token with [d] is somewhat different from the exemplars of "odder". However, the resonance loop linking exemplars on the basis of semantic similarity results in the pattern of activation illustrated in Fig. 6. AQ1

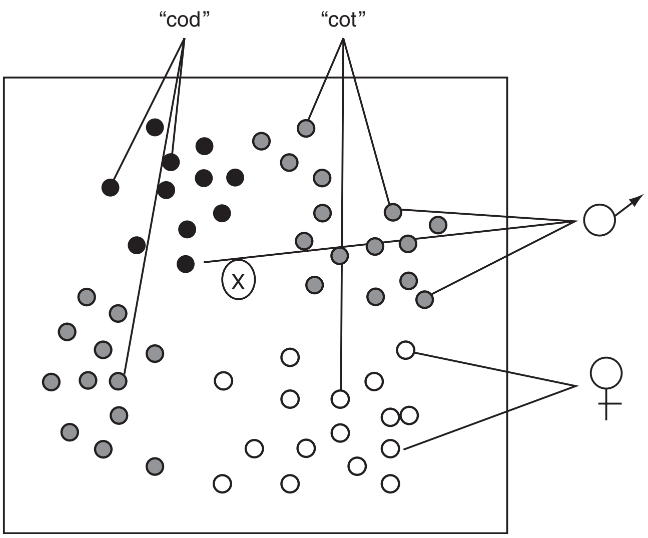

FIGURE 3.5. Illustration that an ambiguous word, indicated with "x", could be interpreted as *cod* more quickly when the speaker is clearly male than when the identity of the speaker is not clear. Each dot stands for an exemplar and the shading of the dot corresponds to the exemplar's activation upon presentation of the "x" token after some resonance. In this case, two categories, "cod" and "male", reinforce each other to categorize the ambiguous token

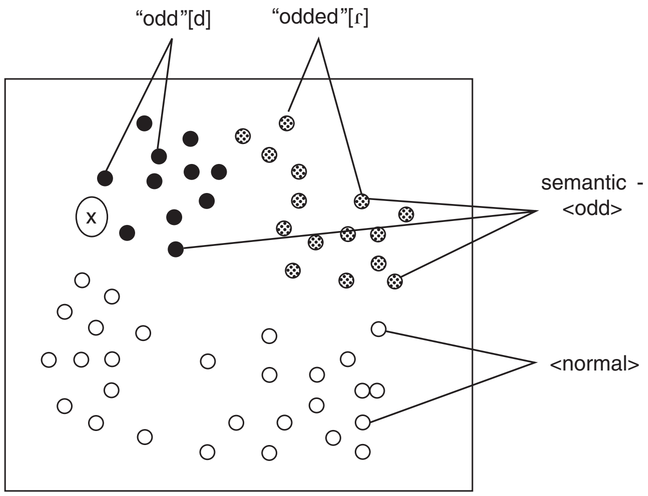

FIGURE 3.6. Illustration of allophonic resonance in the interaction between semantic and phonetic information in exemplar-based phonology. After a new token "x" has been encountered, phonetically similar tokens of "odd" are activated (indicated by darker shading). As a result of semantically guided resonance, the allophonic relationship between [d] and [ɾ] is represented in the pattern of exemplar activation

In this resonance-driven exemplar activation response, the phonological relationship between [d] and [ɾ] is represented as a pattern of exemplar activation in which exemplars of ''odd'' and ''odder'' are co-active despite their phonetic differences.

3.5 CONCLUSION

In this chapter I have suggested that exemplar-based models of phonological knowledge may increase our understanding of the ecology of language—particularly of the cognitive basis of phonological knowledge. I see this project as a continuation of the research aims illustrated so well by John Ohala's work on the phonetic and historical basis of language sound patterns.

One of the main goals of the chapter was to point out that exemplar-based phonology is based on a long tradition of research and theorizing in cognitive psychology and that this general approach to phonological modeling is being pursued by many linguists. There is no one exemplar-based phonology theory, but, rather, a number of nascent models seeking to use this class of memory models to help us better understand, among other things, phonological generalizations and the coexistence of gradience and categoriality in phonological knowledge.

In addition to these general considerations, the chapter outlined my answers to a few important decisions that must be made in exemplar-based modeling of phonological knowledge, suggesting (1) that exemplars in linguistic memory are examples of words from which smaller phonetic/phonological units emerge, and (2) that the representation of exemplars in model simulations should be rich with phonetic detail. The paper also described two important mechanisms in exemplar-based phonology, noting that similarity-matching results in generalization behavior without explicit storage of prototypes and that resonance interactions between phonetic and non-phonetic information in exemplars produces top–down processing influences as well as a representation of linguistically significant sound patterns.

Author Query

AQ1: Please check the word Fig. 6 [There is no Chapter no.]