

# CHAPTER 1

## INTRODUCTION

This dissertation investigates the role of phonological knowledge in speech perception. It proposes a theory of performance which makes use of grammatical competence – specifically, competence expressed in terms of the ranked and violable constraints of phonological Optimality Theory – to weigh competing hypotheses about the phonological structure of the speech signal. This theory is tested empirically against the rival claims of two other models to explain the same phenomena: TRACE, which uses lexical knowledge, and the MERGE transitional-probability theory, which uses segment-string frequency.

The phonological phenomenon with which we are concerned here is phonotactic grammaticality. Languages place tight restrictions on how their segmental inventories can combine into larger units such as syllables, morphemes, or words, and speakers are sensitive to these restrictions in a number of ways. The phonologically systematic patterns of possible and impossible combinations are the phonotactics of the language. Phonotactics causes redundancy, predictability in speech, which the mechanisms of speech perception could in principle exploit.

Phonotactic effects turn up in many places. They appear as systematic gaps in the distribution of sounds in a speech corpus (e.g., Harris 1951, Lamontagne 1993) – what in Chapter 2 are called *phonological gaps*. Phonotactics can drive synchronic phonological alternations, such as that between American English [t] and [ɾ], which are conditioned by the neighboring segments (e.g., Prince & Smolensky 1993). A foreign word can undergo sound changes when it is borrowed that adapt it to the phonotactics of the borrowing language (Itô & Mester, 1994).

Native speakers share intuitions about the phonological grammaticality in their own language of novel phoneme strings (Greenberg & Jenkins 1964, Scholes 1966). English

listeners can also accurately judge the relative frequency of non-English consonant clusters in the languages of the world (Pertz & Bever 1975). A language's phonotactic constraints are respected by its speakers' slips of the tongue (Fromkin 1971) and ear (Sapir 1933; Brown & Hildum 1956; Hallé, Segui, Frauenfelder, & Meunier 1998), and have been shown to influence speakers' perceptions of phonetically ambiguous segments (Massaro & Cohen 1983; Pitt 1998).

In sum, speakers tend to reject phonotactically illegal stimuli in production and perception, requiring, for instance, stronger acoustic evidence to believe that they have heard an illegal stimulus than a legal one. Moreover, illegality measured one way (e.g., off-line intuitive-goodness judgments) tends to agree with illegality measured in other ways (e.g., ambiguous-segment perception). Some sort of language-specific knowledge is being brought to bear on all of these.

What is at issue is the nature of that knowledge, and of its interaction with the mechanisms of language performance. We will be investigating three very different proposals. Each will be examined chiefly in light of its account of the phonotactic effect on ambiguous-phoneme perception, a question directly and explicitly addressed by all three: If a stimulus contains a phoneme which is acoustically ambiguous between one which is legal in that context and one which is illegal, listeners' reports are biased towards the legal interpretation compared to their report of the same ambiguous phoneme presented in a neutral context (Massaro & Cohen 1983).

The claim which I will advance, elaborate, and defend is the following:

- (1) Speech input is parsed prelexically to a featural or phonemic surface representation. When acoustic evidence in the incoming speech stream supports more than one phonological parse, the competing parses are scored with respect to the ranked active constraints of the speaker's grammar, and the more harmonic candidate parse is processed first.

This is a way of allowing performance mechanisms to use linguistic competence, by setting up a perceptual bias against parses which are disfavored by the grammar. It therefore incorporates into the performance theory the traditional linguistic view that the difference between a grammatical and an ungrammatical utterance hinges on whether the utterance fulfills specific formal requirements – whether it meets the structural description of a set of abstract grammatical rules. In this view, an ambiguous phoneme generates two (or more) parses. If one is legal in context and the other is not, perception will favor the legal parse. This is the principal theoretical contribution offered by this dissertation: An account of how phonological grammar can be used in a parsing theory.

Quite different in vision are the two rival theories, TRACE (McClelland & Elman 1986) and the MERGE transitional-probability (TP) theory (Pitt & McQueen 1998). These regard phonotactic illegality as a very concrete phenomenon, equivalent to non-occurrence in the lexicon<sup>1</sup>. Illegality is the extreme low end of a frequency continuum, and its effects are effects of frequency. Where these two theories differ is in how they implement frequency effects.

TRACE is a connectionist model of word recognition, within which phonotactic effects emerge as side effects of the word-recognition process. A stimulus will activate phoneme units, which in turn can produce certain levels of activation in a word unit, depending on the degree to which the stimulus resembles the word represented by that unit. A stimulus containing a phoneme ambiguous between a legal and an illegal one will partially activate some words containing the legal phoneme, but none containing the illegal one. Activation spreading down from the word units to the phoneme units will increase the activation of the unit corresponding to the legal phoneme, which will laterally inhibit the illegal phoneme unit. The result is a perceptual bias towards the legal phoneme.

The MERGE TP theory is associated with a model of phonemic processing, the MERGE model of Norris et al. (2000). In MERGE TP, low-level (pre-lexical) perceptual

---

<sup>1</sup> Transitional probabilities are assigned to a pre-lexical module in MERGE, but the probabilities themselves are computed over the lexicon.

mechanisms keep track of the frequencies with which different phoneme sequences occur. An ambiguous segment and its surrounding context could be interpreted as either of two sequences, but perception will tend to favor the more frequent possibility – that is, it will choose the phoneme that, on the basis of past experience, is more likely in the given context.

Both of these models have demonstrated success in accounting for some of the core phonotactic perceptual phenomena. However, I will argue that neither one is adequate, for reasons crucially connected with their lack of access to abstract grammatical knowledge. Both make predictions that are not borne out, and fail to predict phenomena that occur.

The grammatical approach to phonotactics treats these phonotactic effects as a syndrome, with a single underlying cause, and identifies that cause with listeners' knowledge of the sound pattern of their language in the form of phonotactic constraints against particular combinations of sounds (e.g., Shibatani 1973). Speech tasks make the speaker or listener assign a linguistic parse to the stimulus; parses which are nonexistent or highly marked in the language will naturally be disfavored.

However, most of the "phonotactic" effects are open to another interpretation: Perhaps speakers merely know that some sequences are common and others are rare or nonexistent. The statistical rarity of particular phoneme sequences affects intuitive "possible-word" judgments (Treiman et al. 1996) and ambiguous phoneme perception (Newman et al. 1997; Pitt & McQueen 1998) in much the same way as phonotactic illegality. Rarity also speeds "no" responses in lexical decision and slows same-different judgments of nonwords (Vitevich & Luce 1998). What linguists have described as a categorical contrast between possible and impossible sequences can instead be interpreted as a frequency continuum, with the "impossible" sequences at the irreducible minimum of zero frequency.

Statistical models differ in which statistics they use and how they use them. In the perceptual model TRACE (McClelland & Elman 1986), the rarity of particular sound sequences is encoded in knowledge of words, and statistical knowledge is retrieved by

querying the lexicon. One component of the MERGE model of perceptual decision-making (Norris et al. 2000) keeps track of phoneme-to-phoneme transitional probabilities, which are used without reference to the lexicon. The current version of the Neighborhood Activation Model (Luce 1986; Luce & Pisoni 1988; Vitevich & Luce 1998, 1999) combines knowledge of sublexical sequence frequencies with knowledge of lexical frequencies and neighborhoods. A statistically-based model of the acceptability-judgment task using co-occurrence frequency has been put forward by Frisch, Broe, & Pierrehumbert (1995).

The theoretically most attractive aspect of statistical models is their account of how speakers learn the phonotactics of their language – *en passant*, as a by-product of learning its vocabulary.

On the other hand, they do not explain three phenomena that led people to posit grammars in the first place.

1. If phonotactics is learned simply by learning or hearing words, speakers should be able to acquire any language at all. The lexical and statistical mechanisms only distinguish favored from disfavored sound patterns *within* a language, after the lexicon has been learned or the statistical patterns have been analyzed. Yet English listeners can accurately judge the relative frequency of non-English consonant clusters in the languages of the world (Pertz & Bever 1975). The cross-linguistic commonness or rarity of different classes of segments, sequences, or processes, is not addressed by statistical learning theories, nor is the way in which the processes found in one language resemble those found in others. Capturing these patterns is a central concern of grammatical models of language, which have evolved a wide array of conceptual tools for this purpose: Articulatory grounding of constraints (Archangeli & Pulleyblank 1994), implicational markedness (Greenberg 1964), feature geometry (Clements 1985), natural classes (Chomsky & Halle 1968), and many more.

2. The alternations induced by phonotactics are categorical rather than gradient, and systematic rather than arbitrary. For example, the phonotactics of Standard German forbid

word-final [b d g v z]; in that environment, they turn into [p t k f s], despite their differing frequencies. The frequency difference between (common) word-final [t] and (zero-frequency) word-final [d] is much greater than that between (uncommon) word-final [p] and (zero-frequency) word-final [b], yet German speakers "repair" the illegal final voiced obstruents to the same extent in both cases. And the repair is not to turn the illegal obstruents into the most frequent legal obstruent, but into the *corresponding* legal obstruent.

3. Phonological alternations occur even if the utterance consists of very rare morphemes or even nonce forms, and exceptions to regular patterns are *less* likely to occur as morpheme frequency decreases. These features suggest that the regularity is distinct from the forms it applies to, rather than emergent from them.

The weakness of both non-grammatical theories is the superficiality of their linguistic analysis, which prevents them from abstracting the empirically correct generalizations about legal and illegal sequences.

TRACE and MERGE TP offer extremely simple theories of phonological representation. Phonemes are unstructured lists of features, as in Jakobson et al. (1952). The only phonological domain above the level of the phoneme which is recognized by TRACE is the word, while MERGE TP also recognizes 2- or 3-phoneme sequences. Neither represents the phonotactically crucial domain of the syllable, or any of its constituents such as the onset and rime. Both are incapable of abstracting over features: All patterns are represented at the level of the phoneme, sequence, or word frequencies. More abstract properties which influence phonotactics, such as part of speech or lexical stratum, are not encoded anywhere. The result is that the dependencies these theories represent do not correspond to the ones which are linguistically and perceptually relevant.

These models cannot distinguish *phonological gaps* (sequences which are systematically prohibited) from *lexical gaps* (sequences which are permitted, but missing

from the lexicon through historical accident). If illegality and frequency are the same thing, then zero-frequency sequences should be equally illegal regardless of why they are illegal.

They cannot distinguish *phonotactically relevant context* from *phonotactically irrelevant context*. For example: The nonword [tli] is illegal in English. The illegality of the [l] in that context is due entirely to the context on its left – the word boundary and [t], which create an illegal sequence of two coronal non-continuants in a syllable onset. The [i] has nothing to do with the phonotactic unacceptability of the string; [tli] and [tla] are both illegal. TRACE and MERGE TP are blind to this fact. Each applies a fixed "context" to every phenomenon. The relevant context in TRACE is the entire nonword; that in MERGE TP is the neighboring phonemes. Both theories therefore incorrectly overestimate the perceptual influence of incidental context.

It is an empirical question whether listeners process phonological gaps differently from lexical gaps. Evidence will be presented to show that they can: that phonological gaps are stronger than lexical gaps, and that phonotactically relevant context is more influential than phonotactically irrelevant context.

The organization of the dissertation is as follows:

Chapter 2 discusses the phonological background of the theory – the grammar which it is proposed that performance mechanisms have access to. It first discusses the Optimality-Theoretic approach to phonotactic grammar as a filter on the lexicon, going on to review the distinction between lexical and phonological gaps, and between phonotactically relevant and irrelevant context. Two particularly prominent phonotactic gaps in English syllable onsets – [tl] and [sɹ] – are shown to be phonological rather than lexical gaps, and are analyzed as special cases of more general prohibitions.

Chapter 3 introduces the three theoretical contenders, TRACE, MERGE TP, and the OT grammatical theory. The rationale for each is discussed and the existing empirical evidence weighed, with each theory's account of it presented.

The precise workings of the MERGE TP theory have not yet been explained by its adherents. Much of Chapter 3 is devoted to considering the different design parameters of a theory of transitional probabilities in perception and choosing which possibilities to test. The most important of these parameters is the specific nature of the phonological context: How many segment positions are included, and how do the left and right contexts interact? It will be shown that there is no choice of context that can account for the data cited by the MERGE TP authors in support of the theory. If the context is chosen so as to cover any one part of the data, the theory makes incorrect predictions about the rest. On the chance that some of the contradictory data might be artifactual, two contexts are chosen for testing as the most plausible and interesting.

New empirical evidence bearing on these theories is reported in Chapter 4. The tactical focus is on the distinction between phonotactically relevant and irrelevant context, and on that between phonological and lexical gaps.

Experiments 1-5 build on previous psycholinguistic research on the phonotactic perceptual effect of English syllable structure.

Experiment 1 demonstrates an effect of phonotactically relevant context, but not of irrelevant context, on perception of an [i]–[ɪ] continuum by American English speakers, exploiting the phonotactic illegality of word-final lax vowels. Experiment 2 attempts to replicate Experiment 1 with initial [pw], considered phonotactically illegal by the TRACE authors on statistical grounds (McClelland & Elman 1986), but merely "marginal" by phonologists on the basis of intuition, distribution, and history (Hultzén 1965, Wooley 1970, Catford 1988, Hammond 1999). No effect is found, despite the strong statistical biases against [pw]. Experiment 3 directly compares the bias against [pw] with that against the much more illegal, but statistically very similar, [tl], and finds a much stronger bias against the latter. Manipulations of phonotactically relevant context are found to have the effect predicted by the grammar-based theory, while those of phonotactically irrelevant



context have no effect. These findings are argued to favor the grammar-based processing theory over TRACE and MERGE TP.

Where previous work in this field, including Experiments 1-3, has used stimulus units to measure the dependent and independent variables, Experiments 4 and 5 used a technique which allows the effect of one *response* on another to be measured when judging a CC cluster in which both C's are ambiguous (Nearey 1990). This allows bias effects to be disentangled from stimulus factors and hence measured with greater accuracy. In Experiment 4, the bias against [bw] is compared with that against the much more illegal, but statistically very similar, [dl]. A strong bias against [dl] is found, but none against [bw], corroborating the findings of Experiments 2 and 3. Experiment 5 is a control experiment to insure that the results of Experiments 2, 3, and 4 were not caused by compensation for coarticulation (Mann 1980).

Experiments 6a and 6b exploit the stratified nature of the Japanese lexicon, in which each word belongs to one of four classes with its own syndrome of phonological, morphological, and etymological properties. One of these strata, Sino-Japanese, forbids final [a:], while another, Foreign, permits it. The [a]–[a:] boundary is measured in carrier nonwords containing phonological cues to membership in one stratum or the other.

In Experiment 6b, Sino-Japanese cues are found to bias perception against [a:] as compared to Foreign cues – an effect which is expected and necessary in the grammar-based processing model. The MERGE TP model cannot account for this effect directly, since some of the phonotactically effective context is too far away from the ambiguous segment for the model to capture the dependency. The results can only be accommodated in that model through *ad hoc* revisions.

The phonotactic boundary shift is larger and more robust than a word-superiority effect obtained with the same listeners and paradigm in the control Experiment 6a. This is unexpected under TRACE, which models phonotactic effects as word-superiority effects.

Chapter 5, finally, sums up the claims, arguments, and data presented in earlier chapters, and situates them in the larger research context. Problems and opportunities for the theory of grammar in speech perception are discussed, and areas of future research delineated.