

CHAPTER 5

CONCLUSIONS

Speech perception is guided by the expectation that the stimulus is an utterance in the perceiver's language. This finding cuts across every level of language organization: phoneme inventory (e.g., Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975), phonotactics (e.g., Brown & Hildum, 1956), the lexicon (e.g., Ganong, 1980), and syntax (e.g., Garnes & Bond, 1975).

The preceding chapters have offered an explicit theory of how the mechanisms of speech perception can use grammatical knowledge of the phonology of the stimulus language to arrive at a phonological parse of the input by choosing from a set of candidate parses consistent with the acoustic signal. I have argued that such a theory is necessary if the observed effects of phonology on perception are to be understood, in view of the inability of competing statistically-based theories (TRACE and transitional probabilities) to account for the empirical findings.

Three principal objections were raised against the statistical theories: their inability to distinguish between relevant and irrelevant context, their lack of sufficiently rich phonological structure, and their inability to generalize appropriately to phonological classes. At the root of all three is precisely the feature that makes statistical theories so conceptually attractive – their low degree of abstraction.

The statistical theories, including TRACE and MERGE TP, can be characterized as "unit models" because they attribute perceptual preference for, e.g., [tr] over [tl] to the listener's differing experience of the *specific* phonological units [tr] and [tl]: One is an attested onset and the other is not (Hallé et al. 1998, Pitt 1998), one is common and the other is rare (Massaro & Cohen 1983, Pitt & McQueen 1998), one is supported by many lexical items which contain it and the other is not (McClelland & Elman 1986). The *a priori* plausibility of unit models comes from the pervasiveness of unit-frequency and lexicality effects in language (e.g., Vitevich & Luce 1999, Jusczyk, Luce, & Charles-Luce

1994, Frisch, Large, & Pisoni 2000, Hay, Pierrehumbert, & Beckman in press; Ganong 1980; Samuel 1981, Fox 1984), combined with the minimal nature of the representations they posit – phonemes and words, both of which are needed in any theory. The weakness of unit models is not conceptual but empirical: The phonological knowledge used in speech processing is more complex than can be accommodated in such a simple architecture. The experiments of Chapter 4 were designed to exploit this weakness, in order to argue that a full-fledged phonological competence must be available to perceptual mechanisms on line.

Inability to distinguish relevant from irrelevant phonological context. Phonological processes apply to classes of sounds in classes of environments (e.g., a process of devoicing applying to all obstruents at the end of all syllables). Different rules have different environments. However, the unit models afford only one environment – the word for TRACE, the fixed-length phoneme string for MERGE TP – and are forced to detect linguistically irrelevant accidental correlations involving contextual material which has nothing to do with the actual phonological pattern. Experiments 1–6 showed that in fact, when probabilities are equated, phonologically relevant variation has a much stronger perceptual effect than phonologically irrelevant variation.

This was particularly clear in the case of Experiment 6, where the magnitude of a word-superiority effect was compared directly with that of a stratal phonotactic effect and found to be much smaller. In order to account for the phonotactic effect (which reflected a dependency between the ambiguous phoneme and one three phonemes previous to it), a unit model would have to use such a large environment that it would have to also represent equally strongly the dependency between the first three segments of any word and the fourth – incorrectly predicting a similarly-sized or stronger word-superiority effect.

Lack of sufficiently rich phonological structure. Moreover, since the unit models do not represent syllabification, they could not predict the effects of syllable structure found by Pitt (1998) and in Experiment 5. TRACE, whose phoneme-decision process considers each

phoneme unit in isolation, cannot represent the phonotactic dependency between two phoneme decisions found in Experiments 4 and 5.

Inability to generalize to phonological classes. This problem is acute for MERGE TP, which only represents statistical dependencies between one specific phoneme sequence and another. For this theory, [ki] and [sa] are not two instances of the pattern "CV syllable", but two unrelated phoneme strings. This renders the theory unable to recognize natural classes. Experiments 2–4 indicated that English listeners' experience of the common [labial][labial] onsets [br pr] legitimizes the rare or nonexistent [bw pw], but MERGE TP cannot make the connection. (TRACE's featural level could in principle allow it to capture this generalization, if there were a way of representing syllable structure.) The inability to relate one phoneme string to another exacerbates the problems of irrelevant context and oversimplified phonological structure by preventing the process of comparison which might allow irrelevant factors to be averaged away and lead to the induction of more structured representations.

The OT grammatical theory performed well in all of these tests, predicting shifts when there should have been shifts and no shifts when there should not have been any. The good performance was not due to the specific choice of Optimality Theory – a similar theory could in principle have been constructed around any descriptively adequate grammar – but to the fact that grammatical theory more accurately describes the categories and processes of language. TRACE and MERGE TP both propose, in essence, that the representations and rules active in on-line speech perception are very different from those inferred from typological study of the structure of human languages. Any attempt to elaborate the architecture of either theory to capture more sophisticated linguistic concepts (e.g., by adding a layer of syllable units to TRACE) will amount to building grammar into them. Since their chief conceptual appeal is their promise to explain apparent grammatical effects, like phonotactics, as emergent statistical generalizations captured by a simple learning system, the remedy will require amputation.

There remains some evidence for transitional-probability effects on ambiguous-phoneme perception which are not captured in the OT grammatical theory – the findings of Pitt (1998) on English onset clusters – and the effects of lexicality are thoroughly documented (Ganong 1980; Samuel 1981, Fox 1984, Elman & McClelland 1988, etc.). Given the pervasive nature of frequency and practice effects in all cognitive domains, and their narrow stimulus-specificity (e.g., Klapp et al. 1991, Logan 1988a, 1988b), there can be no doubt that unit-based processes play a prominent role in perception. However, the evidence of this study suggests that they are considerably weaker than the structural constraints of the listener's native language.

Although this dissertation has considered only phonotactic effects of grammar on ambiguous-phoneme perception, the OT grammatical model predicts that effects will be pervasive in more naturalistic on-line tasks. Speech segmentation, for instance, is demonstrably sensitive to the grammar of phonotactics (Norris et al. 1997, Kirk 2001). This can be seen as selection of a grammatically more harmonic prosodic parse over a less harmonic one. Effects of faithfulness should be apparent in word recognition and similarity priming: A nonword which is unfaithful to a real word on a low-ranked faithfulness constraint should activate the word more strongly than a nonword which is unfaithful to it on a high-ranked constraint. Such studies offer a test, not merely of the influence of grammar, but of the specific conception of it put forward by Optimality Theory and competing theories of linguistic competence.

bad for two reasons. 1. prevents emergence of correct complex structural elements by preventing the averaging-out of noise. 2. prevents induction of the classes needed for rules.

weakness of lexical effect understandable -- people hear new words much more often than they hear new syllables or phonotactic violations

TRACE was shown to be inadequate on several grounds: It predicted effects of phonotactically irrelevant context which were not found; it failed to predict a difference in phonotactic strength between illegality and mere low frequency; it predicted a lexical boundary-shift effect to be as large as, or larger than, a phonotactic effect, when the reverse was true.

MERGE TP also failed to account for all of the data. The theory was examined in detail, and it was found that there is no consistent way to set the model parameters so as to explain the data that was offered in support of it. Even when the most plausible of the parameters are chosen, MERGE TP fails to explain our new data. Like TRACE, MERGE TP incorrectly fails to distinguish illegality from low frequency. It also cannot account for the long-distance phonotactic effect observed in the Japanese lexical strata.

The OT grammatical theory performed well in all of these tests, predicting shifts when there should have been shifts and no shifts when there should not have been. There remains some evidence for transitional-probability effects which are not captured in the OT grammatical theory – the findings of Pitt (1998) on English onset clusters.

Although this dissertation has considered only phonotactic effects of grammar on ambiguous-phoneme perception, the OT grammatical model predicts that effects will be pervasive in on-line tasks. Speech segmentation, for instance, is demonstrably sensitive to the grammar of phonotactics (Norris et al. 1997, Kirk 2001). This can be seen as selection of a grammatically more harmonic prosodic parse over a less harmonic one, driven by the

same mechanisms as have been proposed here. Effects of faithfulness should be apparent in word recognition and similarity priming: A nonword which is unfaithful to a real word on a low-ranked faithfulness constraint should activate the word more strongly than a nonword which is unfaithful to it on a high-ranked constraint.

The undoing of both TRACE and MERGE TP is their simplicity. Each has only one phonological domain – the word for TRACE, the fixed-length sequence for MERGE TP – with which it must capture all of the structural generalizations which influence perception. They do not represent such crucial concepts as "onset" or "lexical stratum", which, as we have seen, are perceptually relevant. Any attempt to elaborate the architecture of either theory to capture more sophisticated linguistic concepts will amount to building grammar into them. Since their chief conceptual appeal is their promise to explain apparent grammatical effects, like phonotactics, as emergent statistical generalizations captured by a simple learning system, the remedy will kill the patient.

blah blah blah

SECTION ON EXTENSIONS

The OT grammatical model assumes that the candidate [SR]s include a prosodic parse. In particular, segments are parsed into syllables to the extent that the phonetics of the input and the grammar of the language permit. Where phonetic evidence is inconclusive, the phonetic parser generates multiple candidate [SR]s with different syllabifications, and the usual selection mechanism operates to favor the most harmonic. The motivation for including syllabification in the [SR]s is twofold.

First, syllable boundaries are used in lexical access and speech segmentation. The boundaries may be determined by allophonic cues (Kirk 2000, 2001), by phonotactics

UR = /djuk/	AGREE (VOICE)	ALIGN
a. □ (/djuk/, [vu.djuk])		
b. (/djuk/, [vud.juk])		*

UR = /djuk/	AGREE (VOICE)	ALIGN
a. □ (/djuk/, [vus.djuk])		
b. (/djuk/, [vusd.juk])	*	*
c. (/djuk/, [vu.sdjuk])	*	*