EVIDENCE FOR PHONOLOGICAL GRAMMAR IN SPEECH PERCEPTION

Elliott Moreton
University of Massachusetts, Amherst

ABSTRACT

This paper presents evidence from two experiments 1. that speakers of English use the phonotactics of English in analyzing speech input, and 2. that their phonotactic knowledge takes the form of categorical constraints stated in terms of phonological categories. Experimental results indicate that the phonotactic effect on ambiguous-segment perception is influenced by the relative markedness of the endpoints, but not by their relative frequency.

1. PHONOTACTICS INFLUENCES PERCEPTION: THE MASSARO-COHEN EFFECT

Massaro & Cohen [1] synthesized segments ambiguous between [r] and [l] by varying F3, and asked English speakers to judge them in the following syllabic contexts:

Context	English phonotactics allows
[t^h_i]	[r] only
[s_i]	[l] only
[p^h_i]	both [r] and [l]
[v_i]	neither [r] nor [l]

Table 1. Stimuli used by Massaro and Cohen [1]

Context affected the identification curves: [t^h_i] evoked more [r] responses than [p^h_i] or [v_i], which in turn had more than [s_i]. I.e., a sound acoustically ambiguous between a legal and an illegal segment tended to be reported as the legal one. Listeners demanded stronger-than-normal acoustic evidence for the illegal segment.

One view of this evidence is that a speaker's linguistic competence includes phonotactics, instantiated either as derivational rules [2] or as constraints [3]. I will call this the "grammatical" account of phonotactics.

An alternative is provided by the TRACE theory, a network model of word and phoneme recognition [4, 5]. This model can produce the observed performance effects of phonotactics without any phonotactics being built into the model. Phonotactics, in this view, emerges from lexical statistics: Occurring sequences are facilitated, and non-occurring ones inhibited, by activation spreading from the part of the network which represents lexical knowledge. The attractions of TRACE are that it lets us economize on theory by doing away with a separate phonotactic component, and that it gives a natural explanation for how speakers learn phonotactics.

The TRACE account is, in effect, that when acoustic evidence is inconclusive, the *lexicon* is consulted to see how *probable* each alternative is in the given context, and the decision bound shifts to require stronger acoustic evidence for the *rarer* alternative [4].

The grammatical account is that when acoustic evidence is inconclusive, the *grammar* is consulted to see how *marked* each alternative is in the given context, and the decision bound shifts to require stronger acoustic evidence for the *more marked* alternative.

Since markedness and rarity go together, both accounts make the same predictions in many cases -- but not in all. Under TRACE, the size of the boundary shift should depend on the frequency difference, but not on the markedness difference. Under the grammatical theory, it should depend on the markedness difference, but not on the frequency difference.

We present two experiments in the Massaro-Cohen paradigm which show an effect of markedness difference, but not of frequency difference.

2. EXPERIMENT 1

American English prohibits lax high vowels word-finally. In particular, [i], but not [I], is found word-finally. Hence, both theories predict the [i]-[I] boundary to lie closer to [I] word-finally than word-internally.

Using synthetic two-syllable nonwords, we measured [I]

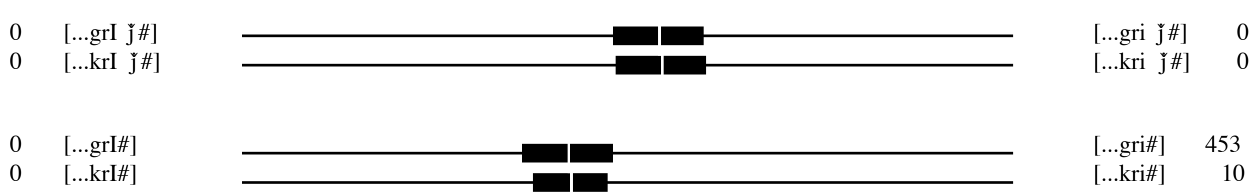

Figure 1. Critical-syllable frequency and subject responses for Experiment 1. Numbers count occurrences of each syllable word-finally per million words. Graphs show mean proportion of [I] responses across the entire continuum, with 0 at the left. Black rectangles are 95% confidence intervals, N=14 Ss.

judgments of sounds on a 5-step [i]-[I] continuum in the contexts [...gr_ǰ #] and [...kr_ǰ #]. Both [i] and [I] are equally rare, though phonotactically legal, in both contexts. The palatal affricate [ǰ] was chosen to minimize coarticulatory influence on the vowel. Under either theory, this should establish a neutral baseline.

To this baseline, we compare [I] responses in the minimally different contexts [...gr_#] and [...kr_#]. In those contexts, [I] is illegal, but [...gri#] is much more frequent than [...kri#]. In fact, [...kri#] is only slightly more frequent than the unattested[...krI#]. Thus, TRACE would predict a larger boundary shift, compared to baseline, in the [g] context than in the [k] context. (Frequency counts are shown along with results in Figure 1).

2.1. Stimuli

The endpoint stimuli were two-syllable nonwords synthesized on the Klatt terminal-analogue synthesizer.

The first syllable of each word was one of [pəl təl səl zəl], filler syllables which were combined factorially with the second syllable. They were chosen to insure that the nonword point came no later than the initial segment of the second syllable.

The second (critical) syllables, which were stressed, are listed in Figure 1.

Intermediate stimuli were made by varying F1, F2, F3, and vowel duration to create a 5-step [i]-[I] scale in the critical syllable.

2.2. Procedures

Subjects were 15 paid student volunteers, monolingual English speakers, tested 4 at a time with headphones in a quiet room. One of them could not hear the endpoint stimuli reliably and was dropped. An AXB classification task was used; Ss judged whether X "sounded more similar" to A or B and responded by button press. Every AXB was also presented as BXA. On each trial, the A and B words were endpoints, differing in only one segment (the vowel of the second syllable for critical trials, the initial consonant for filler trials). The X word was drawn from 5 equally-spaced intermediate steps between A and B. On critical trials, the nonword point occurred at least two segments and 125ms before the critical region.

2.3. Results and discussion

Equally large shifts were obtained for the [g] and [k] contexts, as predicted by the grammatical theory, despite the differences in context frequency. Figure 1 shows mean percent of [I] responses to all intermediate stimuli across the whole continuum. (Frequency counts, per million words of written English, are from the Celex database [5], with the pronunciations Americanized by hand to correct for the original coding of final orthographic *-y* as [I].)

A paired sample *t*-test of the shift between the open and closed syllables showed that a 99% confidence interval excludes 0 ($t_{0.005,\ 14} = 2.977$, $p<0.001$), replicating the Massaro-Cohen effect. The 95% CI for mean difference between the shifts in the [k] and [g] contexts was [-1.66, 2.17], indicating that if a difference exists, it is very small.

This replicates a finding made, with quite different

Figure 2. Critical-region frequencies and subject responses for Experiment 2. Numbers count occurrences of each onset+nucleus word-initially per million words. Graphs show mean proportion of [p] responses across the entire continuum, with 0 at the left. Black rectangles are 95% confidence intervals, N=12 Ss.

stimuli, by Pitt [7]: When markedness is held constant, a change in frequency has no effect.

3. EXPERIMENT 2

The word-initial sequences [t^hl] and [p^hw] are extremely rare in English (see Table 3). There are independent reasons to think that [t^hl] is more marked than [p^hw]:

Distribution: [p^hw pw mw] are marginally attested in loan words with alternate pronunciations like *pueblo, bwana, moiré*; [t^hl tl nl] are not [8]. *Intuition:* The 35 California 7th-graders consulted by Scholes ([9], Experiment 4) found /pwæpθ/ and /bwoitθ/ much more acceptable than /dloipθ/. *Acquisition:* An elicitation study of consonant-cluster errors in English-learning children found that /tw/ is changed to /pw/ more often than /pl/ is changed to /tl/ [10].

Hence, the grammatical theory predicts a larger shift in the [p^h]-[t^h] boundary before [l] than before [w], since the [t^hl] cluster is phonotactically worse than the [p^hw] cluster. The environment [_r], where [p^h] and [t^h] are both legal and of similar frequency, provides a baseline for comparison.

TRACE, however, predicts that the shift size will depend instead on the number (and, presumably, frequency) of words partially activated by the stimulus. This can be manipulated by changing the vowel of the first syllable, as shown in Figure 2.

3.1. Stimuli

The stimuli were two-syllable nonwords with initial stress, consisting of a stop ([p^h], [t^h], or something in between), followed by [r], [l], or [w], followed by one of the nonsense carrier contexts [_Ivnɘm], [_ævnɘm], [_Ifkos], or [_æfkos]. The initial stop was synthetic; the rest was natural speech (mine). Five intermediate stimuli were made by varying F2 in the stop burst and aspiration.

3.2. Procedures

Except for the stimuli, the procedure was identical to that of Experiment 1. Data was collected from 19 Ss. Of these, 12 responded consistently enough to the endpoint stimuli to be analyzed.

3.3. Results and discussion

Results are shown in Figure 2. TRACE predicts that the boundary should shift away from the stimuli beginning with the more frequent strings. Comparison of the [l] and [w] conditions shows that this was not found. The large frequency manipulations scarcely move the boundaries (if anything, the high-frequency strings *attract* the boundary). This replicates Experiment 1's (and Pitt 1998's) lack of a frequency effect.

In two of the four carrier contexts ([_Ivnɘm] and [_ævnɘm]), there was a sizable shift before [l] but none before [w], compared to before [r]. In one ([_Ifkos]) both shifts were of about equal size. These findings are consistent with the grammatical predictions. The differences cannot be explained by lexical statistics, since initial [pw] and [tl] are (about) equally rare in the English lexicon.

(In the fourth carrier ([_æfkos]), the [r] evoked a very high rate of [p] responses -- more even than the [l]. Neither theory predicted this; I take it to be an artifact caused by the abnormally short (10-15ms) [r] in that particular stimulus token. The [r] did not provide a usable baseline, so this part of the data is useless for establishing or refuting an effect of differential markedness.)

4. CONCLUSION

The evidence presented here, along with that of Pitt [8], indicates that, when markedness and and lexical statistics are in conflict, it is markedness than wins out. We have seen that large differences in contextual frequency have no detectable effect on the size of the boundary-shift effect, whether for initial stops or final vowels. Differences in markedness, on the other hand, appear to be modulate the size of the boundary shift in the expected way (though this is far from proven by just this one experiment). This constitutes accumulating evidence that the mechanisms of speech perception have access to phonotactic information independently of the lexicon.

The apparent availability of phonotactics to perceptual mechanisms suggests further that phonological grammar represents phonotactic prohibitions directly, rather than treating them as emergent phenomena of a rule conspiracy. This lends support to constraint-based theories of grammar (e.g., Optimality Theory [3]) over the classical serial-derivational theory of generative phonology [2].

ACKNOWLEDGEMENTS

Many thanks are due to John Kingston and Lyn Frazier for consultation and support. throughout this project.

REFERENCES

[1] Massaro, D. and Cohen, M. 1983. Phonological context in speech perception. *Perception and Psychophysics* 34(4):338-348.

[2] Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English.* Cambridge, MA: MIT Press.

[3] Prince, A. and Smolensky, P. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick: Rutgers Optimality Archive (http://ruccs.rutgers.edu).

[4] McClelland J.L., Elman J. L. 1986. Interactive processes in speech perception: The TRACE model. In D.E. Rumelhart & J.L. McClelland (eds.), *Parallel Distributed Processing*, Vol. 2. Cambridge, MA: MIT Press.

[5] Elman, J. L. and McClelland, J. L. 1988. Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27:143-165.

[6] Baayen, R., Piepenbrock, R., and Gulikers, L. 1995. The CELEX lexical database. (CD-ROM.) Philadelphia: Linguistic Data Consortium.

[7] Pitt, M.A. 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics* 60(6):941-951.

[8] Hultzén, L. S. 1965. Consonant clusters in English. *American Speech*, Feb. 1965:5-19.

[9] Scholes, R. J. 1966. *Phonotactic grammaticality. Janua Linguarum, Series Minor, Nr. L.* The Hague: Mouton & Co.

[10] Smit, A. B. 1993. Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters. *Journal. of Speech & Hearing Res.* 36:931-947.