PHONOTACTIC GRAMMATICALITY IS GRADIENT

Jana Dankovičová (jana.dankovicova@phonetics.ox.ac.uk),
Paula West (paula.west@phonetics.ox.ac.uk),
John Coleman (john.coleman@phonetics.ox.ac.uk),
Andrew Slater (andrew.slater@phonetics.ox.ac.uk)

Oxford University Phonetics Laboratory

INTRODUCTION

Generative phonology claims that English speakers can discriminate between well-formed and ill-formed non-words (e.g. /blɪk/ vs. /bnɪk/) because of their knowledge of phonotactic constraints.

Two experiments were conducted to test the psychological reality of phonotactic constraints. Many such studies have used lexical items and ill-formed non-words as stimuli. In our study, we used only non-words in order to remove the confounding factor of lexical access.

EXPERIMENT 1

Aim

To test whether native English speakers can distinguish between well-formed and ill-formed non-words.

Method

Subjects

6 phonetically naïve monolingual Southern British English speakers. 3 male, 3 female; all right-handed.

Stimuli

84 pairs of well-formed and ill-formed non-words were constructed. Ill-formed stimuli included unattested sequences in onset, nucleus and coda positions, as well as unattested onset-nucleus and onset-coda combinations. The stimuli were randomized and presented to subjects over high-fidelity headphones in a listening booth.

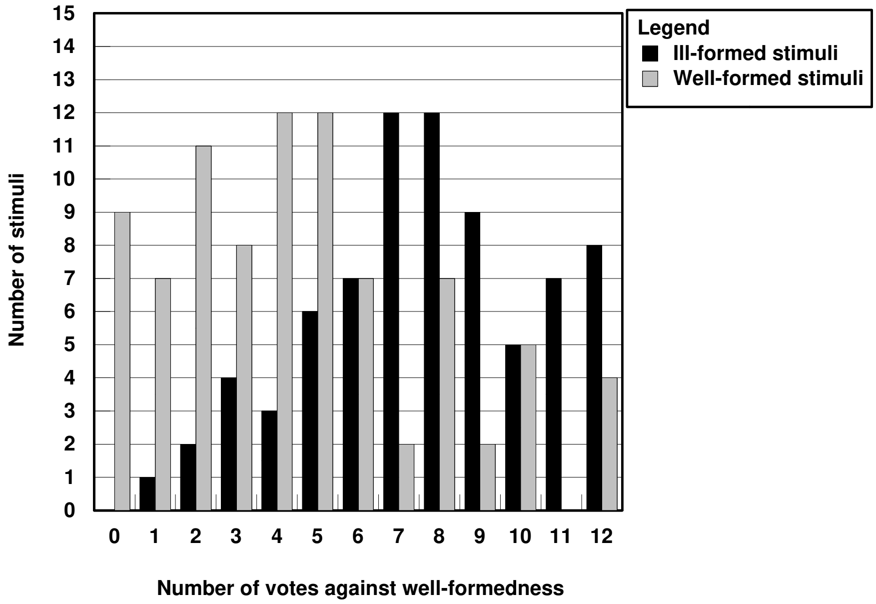
Procedure

Subjects responded by pressing one of two response buttons (‘could be a possible word of English’ or ‘not possible’). A short practise run was followed by 2 runs of the remaining 152 stimuli.

Analysis and results

For each stimulus, the number of responses rejecting it as well-formed was totalled (the ‘number of votes against well-formedness’). Data for all subjects and both runs were pooled (Figure 1).

Figure 1: Votes against well-formedness (Experiment 1)



The distributions of votes for well-formed and for ill-formed non-words were highly significantly distinct ($p < 0.0005$) and ill-formed forms were, in general, judged to be significantly less acceptable than well-formed non-words.

The data contained some surprises; for example:

1) “ill-formed” /mɹuˈpeɪʃn̩/ was more acceptable than some “well-formed” items, e.g. /ˈsplɛtɪˌsɑk /.
2) “well-formed” stimuli composed of low frequency constituents with no close lexical neighbours had many votes against, e.g. /ˈfɪŋkslʌp/ and /ʃəˈlɛnð / which scored 12.

According to Generative Phonology, such cases should not arise at all.

These results, and the distribution of votes in Figure 1, suggested that phonotactic well-formedness is gradient rather than categorical.

EXPERIMENT 2

Aim

To test whether the apparent gradience exhibited in Experiment 1 is confirmed when subjects employ a rating scale instead of a binary choice.

Method

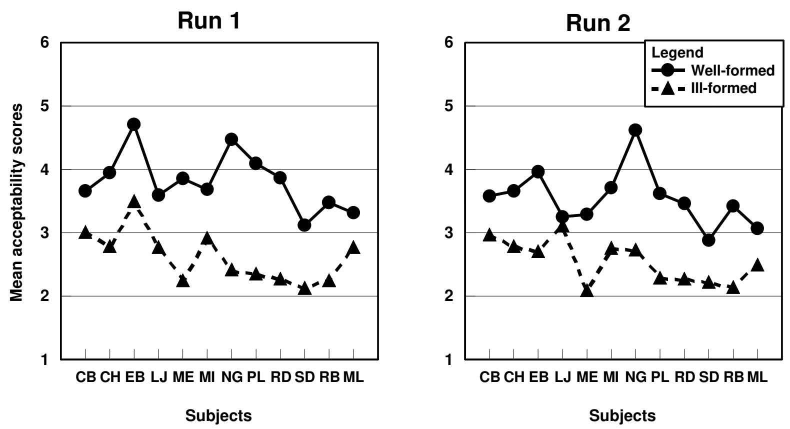
Subjects
12 subjects, (6 male, 6 female); selection criteria as in Experiment 1.

Stimuli and Procedure
As in Experiment 1. Subjects were instructed to rate the acceptability of a word on a 6 point scale (1 = definitely unacceptable; 6 = definitely acceptable).

Analysis and results

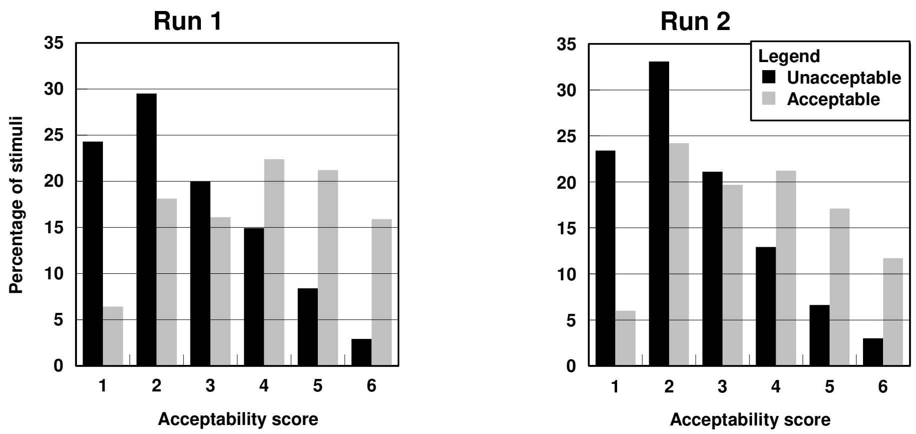
For each subject, mean scores for well-formed and ill-formed stimuli were calculated (Figure 2). The two runs were analysed separately, as 5 out of 12 subjects showed a tendency to judge words less acceptable in run two.

Figure 2: Mean scores for individual subjects (Experiment 2)



Ill-formed stimuli are, in general, less acceptable than well-formed stimuli, for all subjects in both runs.

Figure 3: Distribution of scores (Experiment 2)



The distributions of mean scores (Figure 3) for well-formed and ill-formed stimuli were highly significantly different for both runs ($p < 0.0001$).

All words with onset violations were rated lower than their well-formed counterparts (mean score of 2.3 vs. 4) as were all words with coda violations (mean score of 2.4 vs. 3.4)

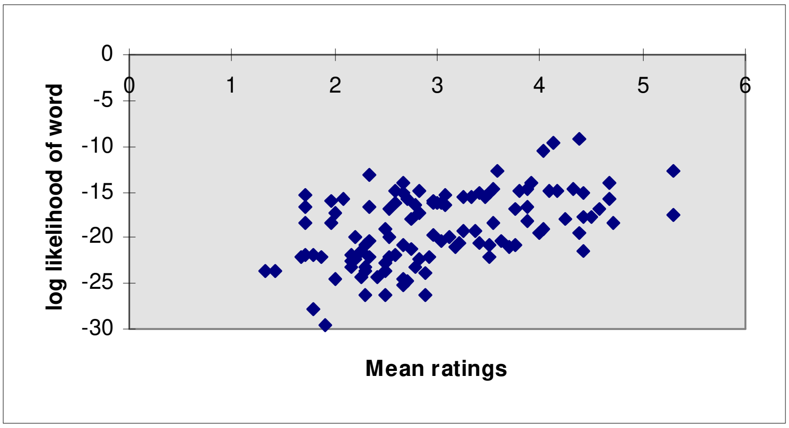
The mean rating of words with onset-coda violations was similar for ill-formed and well-formed stimuli (3.3 and 3.4 respectively)

The 5 words with the worst mean scores had ill-formed onsets. Ill-formed onsets are less acceptable than ill-formed codas, which are less acceptable than unattested onset-coda combinations.

Subjects were fairly conservative in scoring words as acceptable; 62% of "well-formed" stimuli had mean scores < 4 (i.e. unacceptable). Nevertheless, 25% of "ill-formed" stimuli had mean scores > 3 (i.e. acceptable).

In a post-hoc model, we found that the cross-subject mean scores were highly significantly correlated with the log probability of the word ($r^2 = 0.25$, $df = 110$, $p < 0.001$). The word probability was estimated as the product of the observed type probability of the constituent onsets and rimes, as in Coleman and Pierrehumbert (1997). (The probability of ill-formed parts was determined by Good–Turing estimation.) (Figure 4).

Figure 4: Mean ratings against log likelihood of word (Experiment 2)



Conclusion

We interpret the results of the experiments and the model as indicating that although grammaticality can be treated to a first approximation as a binary distinction, a more accurate picture is obtained by regarding it as a gradient (stochastic) property of words.

Reference

Coleman, J. and Pierrehumbert, J. (1997) Stochastic phonological grammars and acceptability. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology.* Madrid. 49–56.

Marslen-Wilson, W. D. and Tyler, L. K. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society of London, B,* **295**, 317-332.