

Modeling the Acceptability of Mandarin Pseudowords with Big Data – the Case of Gradient Phonotactics

Kevin Tang ^{a*}, Kaixuan Gong ^b, Jingyi Yang ^c, Sixian Du ^d and Yanduo Chen ^e

^a *University of Florida*; ^b *University of Cambridge*; ^c *University of Oxford*;

^d *Fudan University*, ^e *Bocconi University*

Motivation: Recent development in psycholinguistics has demonstrated the many benefits of *megastudies* which involve the collection of behavioural data on a large number of linguistic stimuli (Keuleers, 2015). A megastudy can be used to conduct virtual replications of the original experiments in previously published work and to draw new insights by conducting exploratory analyses. In the field of non-word acceptability judgement on Mandarin syllables, there have been a number of auditory lexical decision experiments conducted with different stimuli selection, but they failed to include a complete list of logically possible Mandarin non-words. For instance, Yao and Sharma (2017) conducted their empirical study with only two types of non-words, i.e. tonal gaps and segmental gaps. In the study by Gong and Zhang (2020), only a limited proportion of items were sampled from each non-word type. Therefore, to fill this need in laboratory phonology, we develop a mega database of non-word acceptability judgement that includes all the possible Mandarin non-words. To validate our database, we conducted a virtual replication experiment of Gong and Zhang (2020) which investigated whether gradient phonotactics stems from the violation of different types of grammatical constraints by different non-words (tonal-gap, allophonic restriction, accidental gap and systematic gap) in Mandarin Chinese.

Method: Data: The auditory stimuli in the present experiment followed the permitted order of the Scheme of the Chinese Phonetic Alphabet (Pinyin), which is an authoritative phonetic system that defines clear syllable structure and distinguishes a limited number of each of its components: each stimulus comprised an (optional) onset consonant (22 in Pinyin including zero onset), a (optional) medial glide (4 in Pinyin including zero medial), and a (optional) rime (14 in Pinyin including zero rime) along with a tone (4 in Pinyin), but medial and rime cannot be zero at the same time in one syllable. And the eight syllables with a special structure (single onsets 'm' and 'n' with a tone) were included as some of them are real words. This gave 4848 ($= 22 \times (4 \times 14 - 1) \times 4 + 8$) logically possible Mandarin syllables expressible with Pinyin, among which 3665 non-lexical syllables were found out. 174 low-frequency real syllables (as potential non-lexical syllables) and 15 high-frequency real syllables were blended as control items with the non-words. These 3854 syllables were divided into 25 lists and recorded by a native Mandarin speaker. 349 native speakers of Mandarin Chinese were recruited as participants and divided into 25 groups corresponding to the 25 lists. They were asked to give binary judgement (like a real word or not) on each stimulus they heard as soon as possible, and the reaction time of their NO responses was recorded. **Modeling:** We attempted to model as closely as possible to the original study with a few necessary changes. Instead of ratings, the reaction time (RT) of the NO responses was used, since a non-word with a slow RT can be interpreted as having a high acceptability (Myers, 2015). A rapid NO response can reflect participant's certainty about the

* Corresponding author. Email: tang.kevin@ufl.edu

inexistence of the word. Our syllables were grouped into four types: systematic gaps (missing syllables that violate some principled phonotactic constraint), segmental accidental gaps (missing syllables without any principled phonotactic violations), tonal accidental gaps (existing syllables that cannot bear the high-level tone), and real words. The Pinyin system is unable to represent allophonic variations in vowels, therefore allophonic gaps were not examined. Following the footsteps of Harris, Neasom and Tang (2016) and the original study, an analogical learning model which is distinct from phonotactic learning can serve as a baseline of wordlikeness and allow for a conservative evaluation of the effects of grammaticality of syllable type by first taking the effect of analogical learning into account in a regression model (Albright & Hayes, 2004). A one-edit-distance neighborhood density metric was used as an analogical measure. Stimulus duration was included as an additional control. Numeric variables were log-transformed and centered in the statistical model.

Results: A linear mixed-effects regression model was fitted with the following model structure: $RT \sim type + duration + ND + type:duration + type:ND + (1|item) + (1|Participant)$. The ANOVA table of the model is shown in Table 1. The effect of neighborhood density and duration were replicated in the same directions (ND: β : 0.007519, $p < 0.05$; Duration: β : -0.02717, $p < 0.0001$). Crucially, stimulus type not only significantly improves the model independently, but also accounts for the largest amount of variance in the rating result. The predicted RT of each stimulus type is shown in the figure on the left. The stimulus type follows the same trend as the original study in terms of acceptability (Tonal > Segmental > Systematic), with the exception of the real word condition being out of place. We speculate that this divergence is due to a task effect. In a binary-wordlikeness judgement task, participants would more likely judge real words with a YES response, leaving only a small number of NO responses which in turn leads to a poor estimate of wordlikeness.

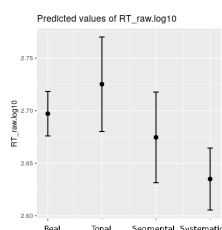


Table 1	Sum Sq.	Mean Sq.	NumDF	DenDF	F value	Pr(>F)
Type	20.1874	6.7291	3	3482.8	91.8796	< 2.2e-16
Duration	8.5762	8.5762	1	3825.9	117.0999	< 2.2e-16
Neighborhood density	0.4575	0.4575	1	3076.0	6.2468	0.012493
Type:Duration	1.1404	0.3801	3	3723.4	5.1905	0.001408
Type:ND	0.4363	0.1454	3	3102.3	1.9858	0.113932

Conclusions: We described a new comprehensive auditory lexical decision database of Mandarin Chinese monosyllabic non-words as a new resource for understanding the mechanism behind grammaticality judgements. Using over 3,800 syllables, a substantial improvement over the 200 syllables originally tested by Gong and Zhang (2020), we were able to replicate much of their key findings: 1) phonotactic acceptability is gradient in Mandarin non-syllables, with the greatest contributor to the gradient being the syllable type; 2) the trend of acceptability is Tonal > Segmental > Systematic. We illustrated that our database can serve as a new resource for phonologists to conduct high power analyses (Coretta, 2020) and to re-evaluate existing claims in Mandarin phonotactic studies. We hope that work will help to guard the field of phonology against a ‘replication crisis’ which is prevalent in psychological sciences.

Key Reference: *Gong, S., & Zhang, J. (2020, May). Gradient Acceptability in Mandarin Nonword Judgment. In *Proceedings of the Annual Meetings on Phonology* (Vol. 8).