Tone-consonant co-occurrence probability in Mandarin

Chiung-Yu Chang, Feng-fan Hsieh

Institute of Linguistics, National Tsing Hua University, Taiwan

Abstract

This study investigates how statistical regularity concerning the distribution of lexical tones and consonantal onsets in Mandarin, formulated as the "tone-consonant co-occurrence probability (TCCP)" here, influences results of wordlikeness judgment task. Native speakers were asked to rate the wordlikeness of monosyllabic real words and pseudowords with existing segmental combinations. Overall, real words with high probability were considered more wordlike than those with low probability. On the other hand, the probability effect was not significant on the well-formedness ratings of the pseudowords. These findings suggest that speakers are sensitive to the tone-consonant co-occurrence patterns, which follow gradual tendencies rather than an "all-or-nothing" manner, but such sensitivity is probably limited to existing forms and cannot be extended to hypothetical ones.

Keywords: phonotactic probability, wordlikeness, consonant-tone interaction, pseudowords, Mandarin

Introduction

It has been long noted that some consonantal onsets rarely co-occur with certain lexical tones in Mandarin Chinese. As a result of the historical tone split, unaspirated obstruent onsets seldom appear on a Tone 2 (mid-rising) syllable, and sonorant onsets rarely occur on a Tone 1 (high-level) syllable (Tung 1970). Nevertheless, there is still a fair amount of counterexamples to these robust tendencies, such as [pa2] 'to pull' and [ma1] 'mother'. Therefore, it is safe to say that the above co-occurrence restrictions are not categorical in nature.

To better capture the probabilistic nature of the tone-consonant co-occurrence restrictions, we propose a new metric that is analogous to conditional phonotactic probability. For a given consonantal onset X and a tone Y, tone-consonant co-occurrence probability (TCCP) for the combination of X∩Y is operationally defined as the type frequency of the co-occurrence of X∩Y divided by the type frequency of X. Further details of calculating TCCP will be provided in the method section.

In addition, we conducted a wordlikeness judgment experiment to test whether native speakers possess the knowledge of Mandarin tone-consonant association. In English, sequence probabilities of segments have been shown to exert unique effects on the wordlikeness of pseudowords after partialling out

neighborhood effects (Bailey and Hahn 2001). By the same token, this study sets out to explore whether TCCP, which involves both segmental and suprasegmental entities, influences wordlikeness in a similar fashion.

Method

Participants

17 native speaker of Taiwanese Mandarin (age range: 20-28, 7 female, without speech or hearing impairment) participated in the experiment. All of them signed an informed consent and were paid for their participation.

Materials

Type frequency was approximated as the number of lexical entries that contain the particular sound or sound combination in the Concise Mandarin Chinese Dictionary (National Languages Committee 2000). The phoneme inventory used for calculation was the same as the phonological representation system adopted by this dictionary, i.e., Mandarin Phonetic Symbols. The resulting TCCPs ranged from 0.002 (for [ʐ]∩Tone 1) to 0.516 (for [k]∩Tone 1). The lowest 14 and highest 15 onset-tone combinations were considered as having low and high TCCPs, respectively.

42 pseudowords with existing segmental combinations (e.g., [tau2]) and 42 real monosyllabic words are chosen as test materials. Half of the pseudowords and real words have a high TCCP, while the other half have a low TCCP. Each item was divided into two or three parts according to the schema — consonantal onset, nucleus, nasal coda (if available), and tone. Neighborhood density (ND) was measured as the sum of the number of monosyllables that have a Levenshtein edit distance of 1 from the stimuli and were entered as a covariate in the statistical analysis.

Procedure

Prompted with visual stimuli written in Mandarin Phonetic Symbols, participants were asked to try producing the (pseudo)words and evaluate their wordlikeness based on a 7-point Likert scale, with 7 being "the most wordlike".

Results

As illustrated in Figure 1, real words with a low TCCP received considerably more low scores than those with a high TCCP. The distinction between high and low TCCPs was not obvious for pseudowords.
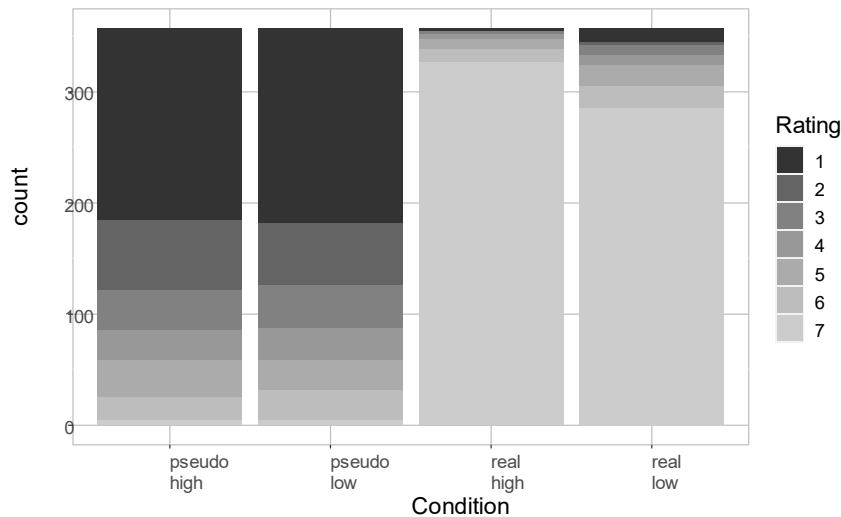
Figure 2. Rating results, where "pseudo" stands for "pseudowords".

The data was analyzed with a series of cumulative link mixed models built by the ordinal package (Christensen 2018) in the R software (R Core Team 2017). All models contained random intercepts for Participants and Items. Lexicality, TCCP, ND, and the interactions among these three factors were entered as fixed effects. Starting from the highest-order term, likelihood ratio tests were performed to assess the contribution of each predictor. The most parsimonious model contained TCCP and its two-way interactions with ND and Lexicality. The significance of each term was confirmed by likelihood ratio tests (all p-values < 0.05).

Separate models were then constructed for the data of real words and pseudowords. Low TCCP was coded as the baseline. Using the method of model selection described above, the best model for real words was the one with only the TCCP effect ($\beta = -1.30$, $SE = 0.53$, $Z = -2.44$, $p = 0.015$), and that for pseudowords incorporated only the ND effect ($\beta = 0.07$, $SE = 0.02$, $Z = 1.91$, $p = 0.056$).

Discussion and conclusion

In this study, we presented a lexical metric to quantify the affinity of specific types of tones and consonants, which helped explain the results of wordlikeness judgment task to a larger extent in the statistic models. The co-occurrence probability is thus a useful method for describing the relationship between consonants and tones.

The significant effect of TCCP on wordlikeness indicates that native speakers are sensitive to the gradient nature of tone-consonant co-occurrence in the lexicon. Interestingly, TCCP interacts with Lexicality and ND. Yao and

Sharma (2017) has reported the influence of ND on the reaction time of real words and pseudowords in a lexical decision task. The fitted regression coefficient of the ND effect observed was generally larger for pseudowords than real words in their experimental results. This is consistent with the current findings that the ND effect was present in the model of pseudowords but not real words. On the other hand, the effect of TCCP was confined to real words, unlike the results in English (Bailey and Hahn 2001). The limited role of TCCP might be attributed to the less significant role of lexical tones in comparison with segments in Mandarin, as evidenced by the lower fidelity of suprasegmental properties in loanword adaptation, as well as many psycholinguistic studies (e.g., Ye and Connine 1999).

A limitation of this study is that only a relatively small set of phoneme combinations are included in the materials. Data collected for more items and from more participants would help us establish a more comprehensive picture of the determinant factors of wordlikeness, including TCCP and other lexical variables.

Acknowledgements

References

Bailey, T.M., Hahn, U. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? Journal of Memory and Language 44, 568-591.

Christensen, R.H.B. 2018. Ordinal - Regression Models for Ordinal Data. R package version 2018.4-19. http://www.cran.r-project.org/package=ordinal/.

National Languages Committee (ed.) 2000. Concise Mandarin Chinese Dictionary. Taipei, Ministry of Education.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Tung, T'.-H. 1970. Hanyu Yinyunxue [Chinese Phonology]. Taipei, Student Book.

Yao, Y., Sharma, B. 2017. What is in the neighborhood of a tonal syllable? Evidence from auditory lexical decision in Mandarin Chinese. In Farrell, P. (eds.) 2017, Proceedings of the Linguistic Society of America 2, 45:1-14, https://doi.org/10.3765/plsa.v2i0.4090

Ye, Y., Connine, C. M. 1999. Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes 14*(5-6), 609-630.