Experimental investigation of Mandarin lexical tone production

Yuyu Zeng[1,2], Robert Fiorentino[1], Jie Zhang[1]

[1]Department of Linguistics, University of Kansas, USA

[2]Department of Communication Sciences and Disorders, University of Wisconsin-Madison, USA

https://doi.org/10.36505/ExLing-2022/13/0051/000593

Abstract

Via two picture naming experiments, we investigated the role of lexical tone in Mandarin spoken word production. In both experiments, target words are monosyllabic, and naming latency is the dependent variable. Experiment 1 (pictures as distractors) examines whether lexical tone is relevant during lexical selection. Trials with exact homophone distractors (zhū vs. zhū) were named significantly more slowly than trials with simple homophone distractors (zhū vs. zhú), suggesting that lexical tone mediates lexical selection. Experiment 2 (pseudo-word transcriptions as distractors) examines the relative timing of lexical tone encoding. It was found that: (1) lexical tone encoding is no later than syllable encoding and precedes the encoding of within-syllable segments; (2) lexical tone encoding takes less time than syllable encoding.

Keywords: spoken word production, lexical tone, Mandarin Chinese, lexical selection, form encoding.

Introduction

The spoken word production process has four identifiable stages: conceptualization, lexical selection, form encoding, and articulation. Much has been learned about the spoken word production process by studying Indo-European languages such as English. Mandarin Chinese, a Sino-Tibetan language, differs from Indo-European languages in many crucial aspects. Research into the Chinese language has yielded valuable insights regarding how language-specific characteristics shape spoken word production. One well-known example is the syllable-retrieval hypothesis of Chinese (O'Seaghdha et al., 2010), which states that form encoding in Chinese spoken word production starts with accessing the syllable, followed by the specification of within-syllable segments; in contrast, in languages such as English, form encoding starts with accessing the segments, which are then assembled into a syllable.

The use of contrastive lexical tone, whose primary acoustic correlate is the pattern of fundamental frequency (F0) realization over a syllable, is one salient characteristic of Chinese. However, little is known regarding the role of lexical tone in Chinese spoken word production. The present study addresses this gap

by examining the role of lexical tone in Mandarin spoken word production, focusing on the lexical selection stage and the form encoding stage.

Experiment 1: lexical tone during lexical selection

Experiment 1 investigates whether lexical tone is relevant during lexical selection in Mandarin spoken word production. If lexical tone is used during lexical selection, exact homophones (zhū vs. zhū) and simple homophones (zhū vs. zhú) should behave differently.

Methods

Thirty-four Mandarin speakers (8 males; mean age = 27.25, SD = 5.52) participated in Experiment 1.

Figure 1 illustrates example items in Experiment 1. The targets are 82 monosyllabic Mandarin words. Each target is paired with an exact homophone distractor, a simple homophone distractor, and two unrelated distractors, corresponding to the three conditions in Experiment 1: Exact Homophone, Simple Homophone, and Unrelated.

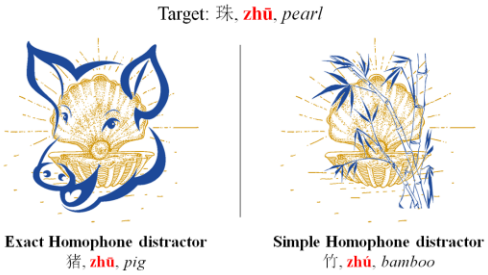

Figure 1. Experiment 1 example items.

Experiment 1 consists of three sessions, which took place over three days over the internet. Session 1 familiarized the participants with the stimuli. Session 2 verified participants' mastery of the stimuli and collected control measurements for statistical modelling. The main picture naming task using picture distractors (Picture-picture Interference) occurred in Session 3, where each unique target-distractor combination was repeated twice.

The naming latency of each trial was manually obtained from the audio recordings of the Picture-picture Interference task. Mixed-effects modelling was conducted using the *lme4* package (version 1.1-27.1) in R (version 4.1.1). The dependent variable was the log-transformed naming latency, and the independent variables include Condition (also included as a random slope) and five control measurements (Sassenhagen & Alday, 2016). The control measurements are not detailed in this paper due to space limitations. The best model was automatically selected using the *buildmer* package (version 2.1). The raw data and analysis scripts can be accessed from this link.

Results

Condition was significant in the selected model (p = 0.011). The Exact Homophone condition (914.60 ms) yielded a longer naming latency than the Simple Homophone condition (908.14 ms; p = 0.061). There was also a significant Condition by Stimulus Repetition interaction (p = 0.038). In

Repetition 1, the Exact Homophone condition (963.39 ms) was named significantly more slowly than the Simple Homophone condition (945.66 ms; $p = 0.001$), but this difference was not significant in Repetition 2 ($p = 0.60$).

Discussion

Experiment 1 revealed evidence for the differentiation of exact homophones and simple homophones, suggesting that lexical tone mediates lexical selection in Mandarin spoken word production. During lexical selection, the competition between exact homophones is stronger than between simple homophones, resulting in longer naming latency in the Exact Homophone condition.

The difference between the Exact Homophone condition and the Simple homophone Condition was significant in Repetition 1 but not in Repetition 2. This is best explained by continuous information flow from lexical selection to form encoding. The phonological overlap between a target and its distractor in the Exact Homophone condition is larger. The facilitation from the form encoding stage (phonological facilitation, henceforth) is thus expected to be stronger in the Exact Homophone condition. In Repetition 1, participants were less familiar with the stimuli, and the phonological activation of the distractors was relatively weak, so the results were dominated by lexical competition. In Repetition 2, participants were more familiar with the stimuli, and the corresponding phonological activation of distractors was higher. Consequently, phonological facilitation overcame lexical competition, leading to no difference between the Exact Homophone and Simple Homophone conditions.

Experiment 2: lexical tone during form encoding

Experiment 2 examines the relative timing of lexical tone encoding by comparing it to syllable encoding, which has been found to precede segmental encoding in previous studies (e.g., O'Seaghdha et al., 2010).

Methods

Twenty-six Mandarin speakers (11 males; mean age = 27.90, SD = 4.84) participated in Experiment 2. The distractors in Experiment 2 are sound transcriptions of Mandarin pseudo-words. This way, the distractors primarily influence the form encoding stage of target naming. The targets are 108 monosyllabic Mandarin words (e.g., 麦, mài, *wheat*), each paired with three types of distractors: (1) Syllable-related (māi); (2) Tone-related (sùn); (3) Unrelated (sún). There are three levels of SOA (stimulus-onset asynchrony) regarding the timing of a target and its distractor: (1) -100 ms, where a distractor precedes its target by 100 ms; (2) 0 ms, where a target and its distractor appear simultaneously; (3) +100, where a target precedes its distractor by 100 ms. The procedure, data processing, and statistical analysis were similar to Experiment 1.

Results

The Syllable-related condition was named significantly faster than the Unrelated condition at all SOA levels (p-values < 0.028). The Tone-related condition (886.43 ms) was named significantly faster than the Unrelated condition (889.21 ms) at the -100 ms SOA level (p = 0.032), and marginally faster than the Unrelated condition at the 0 ms SOA level (921.08 ms vs. 921.74 ms; p = 0.071). The Tone-related condition tended to be named more slowly than the Unrelated condition at the +100 ms SOA level, but the difference was not significant (927.19 ms vs. 924.19 ms; p = 0.10).

Discussion

The Experiment 2 results suggest that lexical tone encoding is no later than syllable encoding, as both Syllable-related and Tone-related conditions were named faster than the Unrelated condition at the -100 ms SOA level. This finding also implies that lexical tone encoding precedes segmental encoding (cf., O'Seaghdha et al., 2010). Moreover, all SOA levels show phonological facilitation in the Syllable-related condition, suggesting that lexical tone encoding takes less time than syllable encoding.

Conclusion

Our study demonstrates that lexical tone participates in lexical selection and form encoding in Mandarin spoken word production. First, lexical tone mediates lexical selection, so the competition between exact homophones is stronger than between simple homophones (Experiment 1). Moreover, after being utilized in lexical selection, lexical tone is also encoded early during form encoding, potentially concurrent with the access of the syllable, followed by the encoding of within-syllable segments (Experiment 2). The Experiment 2 data also suggests lexical tone encoding takes less time than syllable encoding.

Acknowledgments

This work was supported by a Linguistics Research Scholarship awarded to Yuyu Zeng from the Department of Linguistics at the University of Kansas. This work also received support from the Summer Research Scholarship at the University of Kansas.

References

O'Seaghdha, P. G., Chen, J.-Y., Chen, T.-M. 2010. Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. Cognition, 115(2), 282–302.

Sassenhagen, J., Alday, P.M. 2016. A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. Brain and Language, 162, 42–45.