Research Thesis

The Tolerance Principle and the Orthographic-Phonological Mapping of Chinese Characters

Chenyang Lin 0750956

Table Of Contents

1 Introduction

Generalization is an integral process in language acquisition. Normal language development, ranging from lexical acquisition to the acquisition of the form-meaning mappings of morphosyntactic constructions, as well as the acquisition of the semantics of each expression, its associated connotations, and the felicity or appropriateness of its pragmatic usage, would not be possible without the process of generalization, which must start from little to more. However, this naturally leads to the *problem* of generalization, or the problem of *when* the learner should generalize. If generalization is without bounds, then the learner would eventually ignore all linguistic specificities in pursuit of utmost abstraction. Empirically, the risk of overgeneralization indeed poses a challenge for children's acquisition (Marcus et al. 1992). Conversely, if the learner is too conservative about rule induction, then the acquisition would be quite inefficient. How the trade-off is balanced in the minds of every speaker presents a conundrum in language acquisition.

2 Literature Review

2.1 The Tolerance Principle

To solve the dilemma of linguistic generalization, Yang (2016) proposes the *Tolerance Principle* to account for the cognitive "tipping point" between item-based and rule-based learning. He assumes that productive rules are meant to minimize the time complexity (T) of linguistic computation. The establishment of a general rule, however, crucially depends on the number of *exceptions* (e) among the whole set of applicable items (N) that the rule could tolerate. If there turns out to be too many exceptions, then it becomes computationally more efficient to simply retrieve linguistic items on a case-by-case basis. Thus, simplistically, the Tolerance Principle could be *a priori* formulated as follows (Yang 2016:61):

The Tolerance Principle

A rule R is productive if $T(N, e)<T(N, N)$; otherwise R is unproductive.

Based on this assumption, the theoretical threshold (θ) between rule unproductivity and productivity could then be mathematically derived (Yang 2016:64):

R is productive iff $e \leq \theta$ where $\theta := N / \ln N$

Yang's Tolerance Principle provides an intuitive and elegant way to determine and predict when a linguistic rule is inferred. For example, suppose that there are 10 linguistic items to remember in total. At what point does the child decide to generalize the prevailing pattern? By applying the formula, the threshold is calculated to be approximately 4.342, which means that the child would generalize a rule *if and only if* the number of exceptions reaches at most 4. In fact, such categoricity has already been experimentally supported (Schuler, Yang & Newport 2021).

The theoretical postulate of the Tolerance Principle is not without challenges. One criticism is that it does not accurately depict how the mind actually works (Wittenberg & Jackendoff 2018). Indeed, despite its intuitive appeal, it does seem implausible that human cognition checks each lexical entry in a serial manner and keeps them as a list ranked by frequency according to Zipf's Law (1949), as delineated by Yang. This may be resolved by resorting to Marr's (1982) conceptual distinction between the *computational* level and the *algorithmic* or *implementational* level. The fact that, as pointed out by Wittenberg & Jackendoff (2018), the computational level cannot be fully independent from the other levels *does not* automatically refute the conceivability of Yang's theoretical account. Nor does it rule out the possibility that practical considerations may someday be made compatible with the mathematical derivations.

Hernandez, Floyd & Goldberg (2001) conduct an experiment to test the Tolerance Principle. Against the *threshold* view by Yang, they propose the *communicate and access* view, which instead attributes rule productivity to the accessibility of alternative constructions. In the latter view, a rule becomes productive if and only if there is no other linguistic option to convey the same meaning. Their experiment is a within-participants production task, in which children are first exposed to a "mini-language" consisting of novel words mixed with familiar English words (e.g. *lion fep*). After the exposure, the children are asked to label real-world items in the mini-language. Results show that children in the condition with 5 regular and 4 irregular novel words do not generalize the regular word across the board. They argue that their experimental results contradict the hypotheses predicted by the Tolerance Principle, but confirm those predicted by their view. But is this really the case? Their use of artificial stimuli may be problematic. Their artificial stimuli are presented *alongside* English stimuli, which may not prevent pre-established English grammatical rules from interfering with new rule-making. In contrast, Schuler, Yang & Newport's (2021) use of a completely alien language may be more suitable for testing how rule-making develops from scratch. In other words, the conflicting results may simply be due to different language material.

The explanatory power of the Tolerance Principle allows it to account for the learnability of a wide range of linguistic phenomena, including Baker's (1979) paradox concerning English dative shift (Yang & Montrul 2017), the *seemingly* irregular morphological markers of German plurals (Yang 2016:121-136), as well as the ground for the lack of systematicity in Spanish stem alternations, Russian defective inflections, and Polish masculine genitives (Yang 2016:147-156). At the same time, however, the Tolerance Principle has yet to be applied to another important aspect of language acquisition, namely *reading* acquisition. Reading involves several cognitive mechanisms that are quite distinct from speech processing, such as visual graphemic processing and orthographic-phonological conversion (Hillis & Caramazza 1992). By examining the applicability of the Tolerance Principle to the realm of literacy, this study aims to explore the potential of the Tolerance Principle as a more general informational principle that constrains cognition beyond spoken language.

2.2 Writing System

Intuitively, how reading is processed in cognition highly depends on the writing system in question. There exists a rich diversity of writing systems around the world. Daniels (2017) distinguishes six types of writing systems: logosyllabary (representing syllables that differentiate homophony, e.g. Chinese characters), syllabary (e.g. Japanese hiragana), abjad (representing consonants only, e.g. Arabic), alphabetic (e.g. English), abugida (in which some vowels are diacritic, e.g. Sanskrit), and featural (in which phonetic features are imitated, e.g. Korean).

Alternatively, It is possible to categorize all writing systems on the basis of phonographic or morphographic principles (Sproat & Gutkin 2021). Phonographic principles are more *form*-based, which concerns the consistency of the orthographic-phonological mapping. For example, Finnish has a writing system that is fairly regular and transparent in that regard. Morphographic principles, on the other hand, are more *meaning*-based, which concerns the capacity of a writing system to disambiguate homophones. For example, Japanese is a language that abounds with homophones, therefore its writing system would encode the same pronunciation with different symbols depending on the context. Thus the same pronunciation *kisha* could refer to either "reporter" (記者) or "automobile" (汽車). In this study, I will focus on the orthographic-phonological mapping of a given writing system. More specifically, I would like to focus on the orthographic-phonological mapping of *Chinese phono-semantic compounds*, which will later prove to be illuminating for the empirical extension of the Tolerance Principle.

2.3 Chinese Phono-semantic Compounds

The Chinese writing system is built upon *characters* as logographic written units. Traditionally, Chinese characters can be categorized into six types, as in Xu Shen's *Shuowen Jiezi* from the Eastern Han dynasty (Sturgeon 2006): pictographs (象形 *xiàngxíng*), ideographs (指事 *zhǐshì*), compound ideographs (會意 *huìyì*), phono-semantic compounds (形聲 *xíngshēng*), rebuses (假借 *jiǎjiè*) and derivative cognates (轉注 *zhuǎnzhù*). Pictographs mimic the shape of objects (e.g. 日 *rì* "sun"), ideographs designate meaning with abstract signs (e.g. 上 *shàng* "up"), compound ideographs comprise several pictographs or ideographs (e.g. 信 *xìn* "to believe"), and phono-semantic compounds contain both phonological and semantic components (e.g. 江 *jiāng* "river"). On the other hand, rebuses and derivative cognates, though traditionally treated as character types, are technically more like *relations* between different characters. Rebuses group characters similar in form, sound or meaning together (e.g. 考 *kǎo* and 老 *lǎo* "old" are both variations of the same character), while derivative cognates associate pre-existing characters with homophonous words (e.g. the character 長 *cháng* “long” was borrowed from 張 *zhāng* “to stretch").

In particular, phono-semantic compounds are characters analyzable as the combination of *semantic radicals* that hint at the meaning, and *phonetic radicals* that hint at the pronunciation (Liu et al. 2010). For example, the character 河 (*hé* “river”) decomposes into the semantic radical 氵 (*shuǐ* “water”), which implies the aquatic meaning of the character, and the phonetic radical 可 (kě “to be able to”), which implies its *kě*-like pronunciation. It has often been claimed that phono-semantic compounds account for more than 80% of Chinese characters (e.g. Ho, Ng & Ng 2003, Liu et al. 2010), thereby constituting the indispensable backbone of the Chinese writing system.

Studies show that semantic radicals help Chinese character recognition (Feldman & Siok 1999, Chen & Weekes 2004). The contribution of phonetic radicals as phonological cues to visual word processing is also supported by neuroimaging studies (Liu et al. 2020). However, a phonetic radical does not only exert phonological influence, as it has been revealed that phonetic radicals may also elicit semantic priming effects (Zhou & Marslen-Wilson 2002, Tsang et al. 2017). Similarly, semantic radicals do not exclusively activate semantic information either, but show phonological interaction as well (Zhou et al. 2013). Thus even though semantic radicals *primarily* function as semantic cues, and phonetic radicals *primarily* function as phonological cues, their cognitive mechanisms turn out to be more complicated than that.

Ho, Ng & Ng (2003) investigate the developmental trajectory of Chinese schoolchildren’s knowledge of both semantic and phonetic radicals in Hong Kong. With a variety of tasks, it is found that children’s knowledge of radical information correlates significantly with their Chinese word reading ability. Furthermore, a general developmental pattern from a more holistic processing style to a more analytic one also arises. This latter point resonates with Liu et al. (2010), which dissociates the processing of Chinese *words* (which may or may not be composed of more than one character) from that of *characters*. While Chinese words tend to be processed holistically, Chinese characters tend to be processed analytically. This can be shown by how the manipulation of the orthographic arrangement differentially disrupts participants’ performance. Participants suffer the most interference when words in the reverse order also form other words (e.g. 領帶 *lǐngdài* “necktie” vs 帶領 *dàilǐng* “to lead” in reverse), whereas radicals in the reverse order conversely facilitate (non)character decision (e.g. the easily recognizable non-character  vs its real character counterpart 說 shuō “to say”).

While previous studies have shed much light upon the understanding of the psychological workings of Chinese radicals, what remains unknown is whether the orthographic-phonological mapping of Chinese phono-semantic compounds is constrained in any principled way or simply arbitrarily determined. For example, in Ho, Ng & Ng’s (2003) study, their methodology assumed that pseudo-characters were pronounced correctly by children when they pronounced the name of the phonetic radical *or* any character that shared the identical phonetic radical. But this shows that the phonological similarity between phono-semantic compounds and their phonetic radicals is in fact a *fuzzy* concept, allowing for multiple mapping possibilities, ranging from partial likeness to complete identity between their pronunciations (see 2.4). This poses a potential computational challenge for nascent Chinese readers, who must acquire in a sufficiently efficient manner how the pronunciations of a large number of phono-semantic compounds match those of a comparably large number of phonetic radicals to different degrees and in different directions.

2.4 The Tolerance Principle and Chinese Orthographic-Phonological Mapping

What could possibly be the utility of the Tolerance Principle in reading, or more specifically, orthographic-phonological mapping? In more phonographic systems, there seems to be no obvious reason why the Tolerance Principle should have much use, if each orthographic-phonological mapping could have been readily stored without recourse to generalizable rule-making. In more logographic systems, however, the need for rule-making could have emerged due to the relatively vast repertoire of graphemes that a learner must deal with. Chinese is one such example. When Chinese readers read characters, the process is not entirely meaning-based, but the phonological component is always co-activated (Perfetti & Zhang 1995). Indeed, the need for phonological cues to be able to identify a large amount of characters is already reflected in the Chinese writing system itself, in which the majority of characters is phono-semantic compounds. This is the part where the Tolerance Principle could come into play. It is hypothesized that the acquisition of the correlation between phonetic radicals and pronunciations in phono-semantic compounds *also* conforms to the Tolerance Principle, in the sense that the pronunciations implied by the phonetic radicals are individually stored *only if* idiosyncrasies exceed the calculated threshold, and that certain mapping rules between pronunciations and phonetic radicals are generated *only if* otherwise. Moreover, whether a mapping rule is formed or not should be predictable from the proportion of exceptions in the set of characters that contain a given radical.

Inevitably, the preliminary hypothesis that the Chinese orthographic-phonological mapping between phonetic radicals and character pronunciations in phono-semantic compounds is in conformity with the Tolerance Principle requires further refinement. First, what is meant by the "proportion" of exceptions needs to be clarified. It is important that the proportion of exceptions should be measured by *type* frequency, not *token* frequency, since only the former seems to be one of the main driving forces behind rule productivity (Bybee 1995). Thus, orthographic-phonological mapping rules should be dependent on the number of *contexts*, namely the total number of different *words* in which a given phonetic radical occurs, not the absolute number of its occurrence. Second, orthographic-phonological mapping in characters could be further divided into two types: *full* mapping, which maps the whole syllable (e.g. 土 *tu* "earth" and 吐 *tu* "to vomit"); and *partial* mapping, which maps the rhymes only (e.g. 公 *gong* "public" and 訟 *song* "litigation"). Here it is assumed that *both* cases count as mapping rules. Third, although it is *always* the case that any phonetic radical implies some kind of rule based on similarity, the point here is whether the *specific* rule involving the most predominant syllable or rhyme is constrained by the Tolerance Principle. With the phonetic radical 土 as an example, the question is whether the existence of the rule "apply *tu* whenever a character with 土 is encountered" is predictable from the relative weighting of the type frequencies of each character with the given phonetic radical. Finally, the existence of tones also adds complication to the hypothesis. It is possible that phonetic radicals not only constrain the selection of syllables, but also show preference for certain tones. However, since tones could be treated as a separate node of representation from other phonological information (Ye & Connine 1999), this study will only limit its scope to the interaction between phonetic radicals and syllables, whereas tonal information will not be factored in.

To summarize, the research question proposed here is whether the orthographic-phonological mapping between phonetic radicals and pronunciations in Chinese phono-semantic compounds is acquired in accordance with the Tolerance Principle. Crucially, the experiments aimed at the research question should seek to probe the *psychological reality* of the Tolerance Principle, that is, whether the learners' cognition is indeed guided by the Tolerance Principle during reading acquisition. This would be the first study to look into the possible relationship between the Tolerance Principle and reading, which opens a new avenue for research on language acquisition as well as reading in general.

In order to investigate the applicability of the Tolerance Principle in reading, particularly the orthographic-phonological mapping of Chinese phono-semantic compounds, the following two experiments were conducted.

3 Methodology

3.1 Experiment 1

The first experiment tested if the pre-established reading intuitions among Chinese readers were in accordance with the Tolerance Principle. To that end, a pronunciation task was adopted. Proficient Chinese readers in Taiwan were recruited as participants. Since the participants are Taiwanese, traditional Chinese characters were used in this study, thus some discrepancies in orthography and assigned pronunciation with simplified Chinese characters used in other studies are to be expected. Another methodological implication of recruiting Taiwanese participants is that, the type frequencies of the characters as well as subsequent calculations derived from the Tolerance Principle, against which the participants' performance was to be compared, was also based on the Taiwanese corpus, namely the Chinese GigaWord 2 Corpus, the largest traditional Chinese corpus available on Sketch Engine (2017). Finally, the standard pronunciation of the characters also followed that which is officially defined by the Ministry of Education in Taiwan.

In the pronunciation task, participants were asked to judge what they thought the pronunciation of *pseudo-characters* could be. The reason for the choice of pseudo-characters is because real characters cannot tell us whether their associated pronunciations are individually stored or generated by productive rules in the minds of Chinese readers. Only in pseudo-characters, where novel phono-semantic combinations unforeseen in past experience are encountered, can the predictive power of the Tolerance Principle be directly probed. The pseudo-characters were composed by combinations of semantic and phonetic radicals that happen to be accidental gaps in the Chinese writing system, or are infrequent enough that they are unlikely to be readily stored in a typical Chinese reader's mental lexicon.

Although a complete inventory of semantic radicals exists, for example on the website of the National Academy for Educational Research in Taiwan (https://dict.revised.moe.edu.tw/searchR.jsp), to my knowledge no inventory of *phonetic* radicals has ever been compiled, perhaps due to their vastness. Apparently, it would be unrealistic to test every phonetic radical that exists in the Chinese writing system, so in practice only a few were selected as critical stimuli here.

It is also important that the positions of both semantic and phonetic radicals in the pseudo-characters should not violate the conventional regularities of the orthographic structure, otherwise the characters would be seen as *non-characters*, not pseudo-characters (Ho, Ng & Ng 2003). For example, 氵, which only appears on the left side of real characters, should likewise never be anywhere else in pseudo-characters. The goal of the pronunciation task is to let participants judge the *pronunciation* of a character, which always presupposes that the presented characters are *indeed* characters for them to be pronounceable. The blatant violation of fundamental orthographic structures would undermine such purpose.

The pronunciation of pseudo-characters by the participants was then compared with that predicted by the Tolerance Principle. If the Tolerance Principle is indeed operative in the orthographic-phonological mappings of phono-semantic compounds, then the pronunciation should correspond to the pronunciation of real characters with the given phonetic radical that have the highest type frequency *only if* the type frequency crosses a certain threshold, and show other possibilities *only if* otherwise. The type frequency is calculated by the occurrence of words in which characters containing the given phonetic radical take part. As a consequence, a distinction is drawn between word and character (Liu et al. 2010).

3.1.1 Materials

For Experiment 1, 20 phonetic radicals were handpicked from the Chinese GigaWord 2 Corpus. For each phonetic radical, the type frequencies of characters with the given phonetic radical were catalogued. The type frequency of each character was operationalized as the total number of items with the given character in the corpus, which served as the proxy for the variety of real contexts or words in which the character (and by extension the phonetic radical as its subcomponent) can occur. Only characters with at least one item in the corpus were considered, since it is unknown whether characters that are not attested in the corpus circulate in general usage at all. It may be argued that one could go further, taking only characters with the most dominant type frequencies into account. However, there are several reasons for the inclusion of minority characters as far as the corpus allows. First, the Tolerance Principle presupposes a *categorical* distinction between item-based and rule-based learning, thus it is possible for any character even with the type frequency of one to tilt the scale in its favor. Second, it is unclear where to draw the line between numbers of type frequencies considered and those that are not without arbitrariness. Finally, if any Chinese reader happens to know any rare character with a given phonetic radical, then its effect on the rule-making process concerning the phonetic radical should not be easily ruled out.

The type frequencies of each character served as background parameters for the calculations of the theoretical thresholds for each phonetic radical as predicted by the Tolerance Principle (see 3.1.3). The selection of the phonetic radicals was carefully designed in such a way that, among all phonetic radicals, 10 fell under their thresholds, while the other 10 passed their thresholds. This facilitates direct comparison between the *above-threshold* condition and the *below-threshold* condition.

A pseudo-character was assigned to each phonetic radical, which was operationally defined as any character that does not have any lexical entry in the corpus. Thus even though the selected pseudo-characters technically *do* have real pronunciation (as designated in official dictionaries), it is assumed that their occurrences are so rare, to the extent that there is no instance even on a large corpus-wide scale, that these characters are highly unlikely to be readily entrenched in the minds of normal Chinese readers. In the experiment, participants only saw the pseudo-characters as their stimuli, so their orthographic similarity with other real characters was irrelevant. However, to avoid the possible interstimulus priming effect of semantic radicals (cf. Feldman & Siok 1999, Chen & Weekes 2004), it was ensured that no semantic radical was repeated among all 20 pseudo-characters.

The full list of phonetic radicals and their respective parameters and associated pseudo-characters was uploaded on Yoda, an online repository in Utrecht University.

3.1.2 Participants

The participants were all Taiwanese, recruited via the social network of the experimenter. Participants that did not fall between the age of 18 and 69, as well as participants with dyslexia, were automatically excluded from participation. There were 28 participants in total (19 men and 9 women), with the mean age of 22. The participants did not receive any compensation for the experiment.

3.1.3 Procedure

The online experiment was designed with jsPsych (de Leeuw, Gilbert & Luchterhandt 2023), an open source JavaScript framework for online behavioral experiments. The link to the experiment was directly sent to participants. Participants may finish the experiment on any web browser with their own computers or tablets, but not on smartphones. It began with the informed consent asking for voluntary participation, followed by a survey that collected personal information about age, native language, multilingualism, dyslexia, biological sex, and handedness. Then, the participants were asked to activate their built-in microphone for recording. There were 5 seconds for them to test their microphone, followed by a playback bar for them to check whether they could hear themselves clearly, and if not, they had the chance to retry the test recording. Afterwards, the instruction was presented, telling the participants that they were about to see some Chinese characters consecutively, and that they should say what they thought the pronunciations were by intuition within the time limit. Furthermore, they should also try to articulate one pronunciation only for every character. In the experiment, there were 2 practice trials and 20 experimental trials in total. Each trial lasted 3 seconds, with 1 second of rest between trials that showed a fixation cross. In each trial, the participants saw the display of a particular Chinese pseudo-character along with a timer to let participants know how much time was left, during which they must record their response. The imposed time limit forces participants to really draw on their reading intuition, and also prevents cheating (e.g. looking up in a dictionary or googling). The trials continued non-stop regardless of how participants responded. All participants saw the same set of experimental items in a different randomized order.

The audio recordings of the participants were stored in .wav or .webm format depending on the used device, along with .csv files that traced the exact sequences of pseudo-characters encountered by each participant. With these combined, it was then possible to manually transcribe the audio recordings into *pinyin* in the .csv files, matching each recording with its associated pseudo-character. Recordings that matched the dominant pronunciation for a given phonetic radical were labelled as 1, those that matched the *non*-dominant pronunciation for a given phonetic radical were labelled as 0, and those that did not match any pronunciation, as well as those whose audio quality was unintelligible, were labelled as NA. By doing so, the participants' actual performance could be quantifiably compared with the theoretical predictions based on the Tolerance Principle.

Here is an illustration of how the comparison works. Suppose that 匀 (*yun*) is the targeted phonetic radical. When combined with the semantic radical 虫 (*hui*, "insect"), a pseudo-character 虳 can be generated. It is then possible to search for the total occurrences of words with characters that contain 匀 in the Chinese GigaWord 2 Corpus. According to the corpus, the type frequency is 719 for words with 均 (*jun* "even"), 791 for words with 鈞 (*jun* "an ancient unit for weight"), 88 for words with 匀 as a standalone character (*yun* "even"), and relatively negligible for the remaining characters. Based on these figures, we get both θ and N, which tell us whether the assigned pronunciation for the pseudo-character should follow the rule of the many or not. N is simply the sum of all type frequencies, which is calculated to be 1614, whereas θ is the division of N by the natural logarithm of N, which equals approximately 218.51. This final number predicts that an orthographic-phonological mapping rule for the dominant pronunciation, which is *jun* in this case, is productive *only if* the type frequencies for the exceptions, namely whatever pronunciations that are different from the dominant pronunciation (e.g. *yun*), *do not* exceed that number. For the example at hand, not only does the pronunciation *jun* have the majority vote, but the threshold has also not been breached by alternative pronunciations. Consequently, the orthographic-phonological mapping rule between the pronunciation *jun* and 匀-containing characters is predicted to be *productive*, in the sense that it is transferable to pseudo-characters, as shown by the more consistent judgment of the pronunciation of 蚐 across participants. Theoretically, strict consistency would require that all participants choose the dominant pronunciation *jun without exception*, however the more realistic expectation would be that the proportion of the pronunciation judgment of 蚐 as *jun* is significantly higher than the proportion of the pronunciation judgment of any character in the below-threshold condition as its dominant pronunciation.

Here is another illustration for phonetic radicals whose orthographic-phonological rules are predicted to be *unproductive* by the Tolerance Principle. The phonetic radical 賣 (*mai*) has the total type frequency (N) of 2672 and the theoretical threshold (θ) at 338.63. The latter number is surpassed by the type frequency of 續 (*xu*, "to continue," 811) as well as that of 讀 (*du*, "to read" 545). This means that, no matter how predominant the majority pronunciation is (*mai*, 1001), the Tolerance Principle predicts that *no* robust orthographic-phonological rule is formed for 賣. Instead, it predicts that the learning of pronunciations associated with 賣 is *exemplar*-based, as reflected by the inconsistent judgment across the participants. It is expected that such inconsistency would translate into a not significantly high proportion of the dominant pronunciation *mai* in the pronunciation judgment of the pseudo-character 韇, in comparison with the proportion of the dominant pronunciation of any character in the above-threshold condition.

The illustrations above outline the way whereby the theoretical prediction for each phonetic radical could be empirically tested. By comparing the mathematically derived predictions with the actual results produced by participants, it becomes feasible to directly examine the possible role of the Tolerance Principle in reading Chinese characters.

The experiment had been approved by the Faculty Ethics Assessment Committee Humanities (FetC-H) of Utrecht University, with the reference number 23-058-02.

3.1.4 Results

The dataset used for statistical analysis was uploaded on Yoda. For the generalized condition, the expected probability was always 1, whereas for the not generalized condition, the expected probability was simply the proportion of its type frequency in the corpus. The number of experimental trials multiplied by the number of participants produced 560 data points in the whole dataset. All data points with NAs were then removed from the dataset, resulting in the remaining 496 data points for subsequent analysis. The simple count of 1s and 0s in the remaining raw data indicated that, as predicted, the generalized condition indeed elicited a higher proportion of dominant pronunciations (140 : 96) in comparison with the not generalized condition (101 : 159), illustrated in Figure 1.

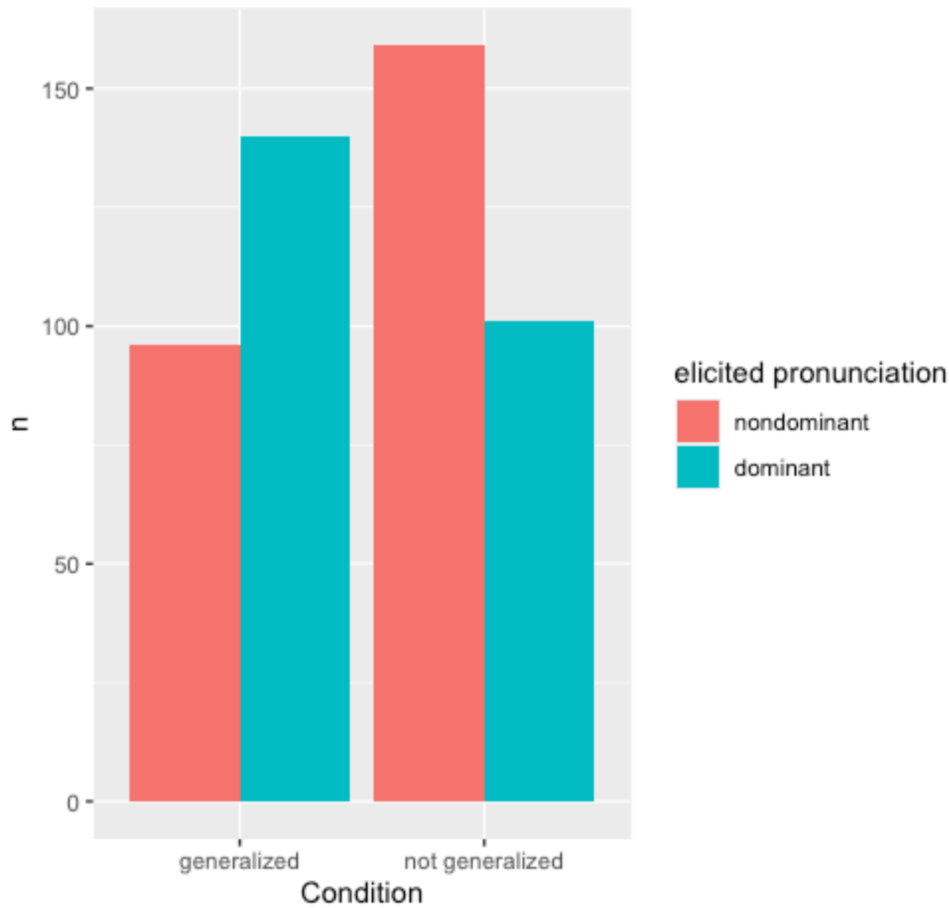

Figure 1: Number of Dominant/Non-dominant Pronunciations Per Condition

To test whether such difference between both conditions was statistically significant, further statistical analysis was run by R (R Core Team 2022). More specifically, the *glmer* function from the *lme4* package (Bates, Maechler, Bolker & Walker 2015) was used to construct binomial generalized linear mixed models to analyze if the experimental conditions (generalized vs not generalized) were able to explain participants' differential judgment of the pseudo-characters.

The intercept-only model, in which only the random intercepts of different participants and stimuli were factored in, produced the warning message of singularity, possibly because the between-participant variance was near zero, yet was treated as exactly zero by the algorithm due to constrained optimization (Bates, Maechler, Bolker & Walker 2015). Visual inspection of the data proved that, in comparison with between-stimuli variance, between-participant variance was indeed quite imperceptible, as shown in Figure 2.

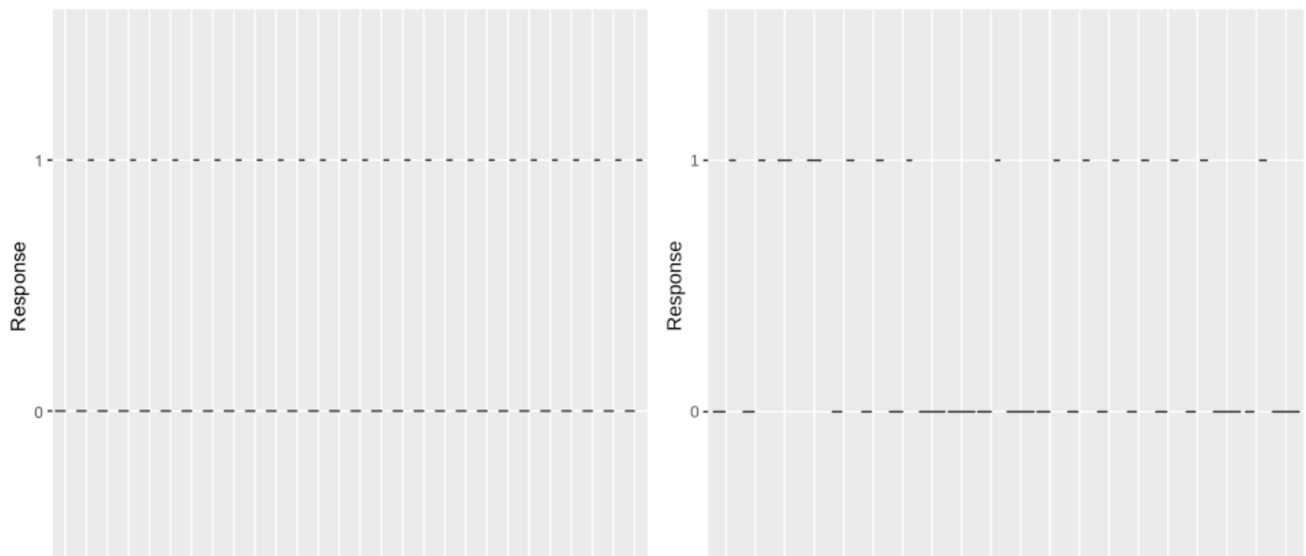

Figure 2 Violin Plots of the Variability between Subject and Stimuli

In order to fit the intercept-only model, the random intercept of participants was removed. Ideally, the random effects should be kept maximal (Barr, Levy, Scheepers & Tily 2013), but considering that the between-participant variance was estimated as zero, its removal should not affect the statistical results too much, albeit with the risk of oversimplification. The simplified model, with stimuli as the only random intercept, did not encounter the problem with convergence.

A mixed model with the experimental conditions as the fixed effect and the stimuli as the random effect was then compared against the previous model to see if the experimental conditions accounted for extra variance of the data. Although the mixed model showed that the not generalized condition was indeed negatively correlated with the occurrence of dominant pronunciations (-0.706), the ANOVA test showed that the difference between both models was not statistically significant ($p=0.2463$), suggesting that the experimental conditions, defined entirely by being situated either above or below the theoretical thresholds derived from the Tolerance Principle, were not particularly predictive of participants' responses.

For each item, the probability of success predicted by the mixed model was further compared with the *expected* probability of success, the latter of which was equal to the new "threshold" variable. The threshold variable was simply the mean of threshold values for both experimental conditions. Thus, the threshold value of all items in the generalized condition was 1, while the threshold value of all items in the not generalized condition was 0.6026139. The difference between the modeled probability of success and the expected probability of success was then factored in a new mixed model, which tested the possible interaction between the experimental conditions and the difference. If there was an interaction effect, then this would imply that experimental conditions did in fact bias the responses, provided that the threshold values were taken into consideration. However, the model showed that the interaction effect was again not significant ($p=0.791$), corroborating results from the first mixed model. Furthermore, the model did not successfully converge.

Finally, a two-sided one-sample proportion test was conducted to check if the actual proportion of dominant pronunciations in the not generalized condition[1] (i.e. 101:159) was sig-

[1] The same test could not be done for the generalized condition, since the expected proportion of dominant pronunciations in that case would be 1 and thus incomparable.

nificantly different from the expected proportion of dominant pronunciations, which was simply the mean threshold value mentioned above (i.e. 0.6026139). If the null hypothesis that the actual proportion was no different from the expected proportion turned out to be true, then this would mean that type frequency as a factor was already sufficient for the prediction of the actual proportion. On the contrary, results showed that their difference was indeed quite significant (p=0.00000000000269), which meant that the type frequency of dominant pronunciations alone was unable to account for the actual proportion of dominant pronunciations in the not generalized condition.

3.2 Experiment 2

The second experiment tested if the orthographic-phonological mapping rules of Chinese characters could also be acquired by non-Chinese readers. Since proficient Chinese readers are already biased by their prior knowledge of the writing system, the selection of *non*-Chinese readers could serve to simulate the acquisitional process that incipient Chinese readers would have undergone during their initial exposure to characters. In other words, while Experiment 1 tested if the Tolerance Principle was already operative in the minds of Chinese readers, Experiment 2 tested if the *acquisition* of Chinese orthographic-phonological mappings was also constrained by the Tolerance Principle.

This experiment drew much inspiration from Schuler, Yang & Newport (2021). In their study, two experiments were carried out to investigate whether the acquisition of an artificial morphological rule (i.e. adding a plural marker *ka*) by children and adults followed the Tolerance Principle. In both experiments, the participants were divided into two conditions: the above-threshold condition and the below-threshold condition. The threshold was determined by the Tolerance Principle, which was approximately 4.096 in the case of 9 distinct items, thus the above-threshold condition consisted of 5 regular and 4 irregular items, while the below-threshold condition consisted of 3 regular and 6 irregular items. During both experiments, participants were first exposed to the artificial stimuli matched with pictures (e.g. *gentif mawg ka* for plural objects), and then underwent a production test similar to the wug test (Berko 1958) to test whether a productive rule was formed, as well as a rating test to gauge which set of items participants actually learned from the exposure. Results from the first experiment showed that children's acquisition indeed followed the Tolerance Principle, while adults' did not. Results from the second experiment, in which the distribution of items was intentionally designed to be closer to natural language, also showed that only children's acquisition followed the Tolerance Principle, but only when each child's individual-dependent inventory of items were taken into account. In the spirit of Schuler, Yang & Newport (2021), the experimental design here will re-semble theirs in many respects, albeit not without important modifications for present purposes that will be elaborated upon in detail in the following sections.

3.2.1 Materials

The experimental materials for this experiment were partly derived from Experiment 1. Since a list of Chinese phono-semantic compounds has already been made in Experiment 1, and that non-Chinese readers are normally unfamiliar with Chinese characters, there was no need to generate artificial stimuli here. For present purposes, only *one* set of real characters would suffice for the testing of new rule formation, which, like Experiment 1, was characterized by a shared phonetic radical 土 (*tu*). However, in this case, the background parameters (i.e. the real characters) came to the fore as experimental stimuli as well. Thus, in addition to 土, real characters such as 吐, which functioned merely as part of the behind-the-scenes calculation for the Tolerance Principle in Experiment 1, now served as part of the exposure stimuli that participants saw as well (see Table 1 below), which were intended to train the participants to attune themselves to the associations between the phonetic radical and certain pronunciations. Another difference from Experiment 1 was that, even though the exposure stimuli of this experiment were real characters, their *type frequencies* were artificially proportioned. Since non-Chinese readers have no prior experience in reading Chinese, they would remain insensitive to the real type frequencies of real characters in the world of texts. This would open up the opportunity for the experiment to redefine the type frequency of each real character on its own terms, operationalized as the occurrences of the real character *within* the experiment, during which non-Chinese readers see Chinese characters for the first time. In other words, the connection between *real* type frequencies and *experimental* type frequencies of real characters was severed in this experiment.

Nonetheless, in order to replicate the ecological validity of the experimental stimuli (cf. Experiment 2 in Schuler, Yang & Newport 2021), the assigned experimental type frequencies followed the Zipfian distribution, as is the case in natural language. Depending on the redefined, "pseudo" type frequencies, the *above-threshold* condition and the *below-threshold* condition were created. In the first condition, the pseudo type frequencies of the real characters were proportioned such that the type frequency of the dominant pronunciation exceeded the theoretical threshold as calculated from the Tolerance Principle, whereas, in the second condition, the pseudo type frequencies of the real characters were proportioned such that the dominant pronunciation fell under the theoretical threshold. It is worth emphasizing that the pseudo type frequencies of the real characters do not have to be proportional to their real type frequencies. Indeed, the point was to observe whether the Tolerance Principle would still work even with these radically modified type frequencies. Moreover, to ensure that the pronunciations of the real characters match the specifications of both conditions, the pronunciations of some real characters were replaced with the dominant pronunciation instead. For the naive non-Chinese reader, the experimentally defined mapping would be as arbitrary as the grammatical one. Thus, although not all real characters retained their original pronunciations in the experiment, this should not affect the new rule formation of non-Chinese readers.

A final note is that the real characters in this experiment always occurred in word-forming collocations. This means that, in effect, *words* would be the basic units of the experimental stimuli, not *characters* (cf. Liu et al. 2010). Indeed, characters in real texts seldom show up in isolation, but are always contextually embedded in a myriad of different words. Without the collocational dependencies, it becomes somewhat paradoxical to define the *type* frequency of characters, since the repetition of a character in itself could only increase its *token* frequency. To this end, 160 unique real words were chosen, each representing a combinatorial possibility of a character. There were 7 targeted real characters in total, each co-occurring with a subset of words. These 7 subsets conformed to the Zipfian distribution (e.g. the largest subset was twice as large as the second largest, etc.). Although non-Chinese readers would not be able to guess the meaning behind these words, it is conjectured that the presentation of the collocational possibilities alone is sufficient for them to become sensitized to the underlying orthographic-phonological tendencies to the extent that a rule is formed, albeit implicitly.

The entire set of characters and the proportioned pseudo type frequencies for each real character is shown in Table 1. With regard to experimental conditions, given that the total number of targeted real characters is 7, the Tolerance Principle predicts that participants will generalize the dominant *tu* pronunciation in the case of 4 dominant and 3 non-dominant items (4R3E condition), but will not generalize it in the case of 3 dominant and 4 non-dominant items (3R4E condition). The small difference in frequency distribution between both conditions puts the *categoricity* assumed by the Tolerance Principle on trial. Indeed, as the logical consequence of the Tolerance Principle, it seems like a strong prediction that merely 15 out of 160 items would have such disproportionate impact on the formation of a general rule.

Type Frequency	Character	4R3E Condition	3R4E Condition
60	土 tu	tu	tu
30	吐 tu	tu	tu
20	杜 du	tu	tu
15	肚 du	tu	du
15	社 she	she	she
10	牡 mu	mu	mu
10	灶 zao	zao	zao
160			

Table 1: Stimuli of Experiment 2

3.2.2 Participants

Although, as shown by Schuler, Yang & Newport (2021), only children are prone to categorical learning, while adults are prone to probability matching, this experiment still recruited adults as participants, which aimed at the reproducibility of Schuler, Yang & Newport (2021)'s finding concerning adults. In other words, this experiment tested whether acquisition among adults was in accordance with the Tolerance Principle, or instead in a more probabilistic fashion. Admittedly, it would be quite impossible to regard non-Chinese readers as genuine “blank slates” without interference from other languages, but this could still be one viable way to assess the *generalizability* of the role of the Tolerance Principle in the acquisition of Chinese orthographic-phonological mapping, irrespective of the sociocultural backdrop of learners.

Given that, as shown in Table 1, the phonological distinction between the aspirated *tu* and the unaspirated *du* would be indispensable for the acquisition of correct orthographic-phonological mapping and subsequent judgment, participants must be those that already internalized such distinction. In other words, participants must already speak another language that similarly has phonemic aspiration for the acquisitional process to work. Coincidentally, the Hindustani language, which includes both Hindi and Urdu, is a language that manifests such a property. Thus, for this experiment, Hindustani-speaking participants were recruited via the social network of a fellow student at Utrecht University, as well as local networks in Taiwan. It is worth noting that, even though the mastery of Hindustani was set as the criterion, it need not matter whether the participant spoke it as his/her native language, or merely spoke it as a lingua franca. The point was simply that the participants were sensitive to the linguistic property of phonemic aspiration present in the Hindustani language. In the end, a total number of 22 Hindustani-speaking participants participated in the experiment.

3.2.3 Procedure

Like Experiment 1, Experiment 2 was also designed with jsPsych. The experimental stimuli consisted of the visual display of Chinese words accompanied by their recorded pronunciations. The audio recordings were recorded in the phonetics lab of the Institute for Language Sciences (ILS) Labs in Utrecht University. The Sennheiser ME-64 microphone in the soundproof cabin picked up the recordings, which were amplified by the Symetrix 302 microphone preamplifier. The recordings were processed with the software Audacity (Audacity Team 2014), which exported them as .wav files for further editing. Since participants’ response time was not of interest here, the duration of each audio file need not be uniformly controlled for, as long as the overall integrity of each audio file was not compromised. Still, all audio files were cut into around 1 second, and were uploaded on Yoda.

The participants were randomly assigned to either the *above-threshold* group or the *below-threshold* group, who saw separate lists of stimuli. As can be seen in Table 1, both lists differ only in the pronunciation of 肚 (*tu* vs *du*), which "tips" the relative weighting of all pronunciations. It was ensured that the number of participants in both groups was counterbalanced. This entails that, unlike Experiment 1, Experiment 2 adopted a *between*-subjects design.

Before the exposure phase, the participants first went through the information letter. Next, as in Experiment 1, participants were asked questions about personal information, including age, native language, multilingualism, dyslexia, biological sex, and handedness. Importantly, given that this experiment investigated the *acquisition* of the orthographic-phonological mapping rules of characters, participants were additionally asked if they spoke Hindi, and if they were already able to read Chinese or Japanese*.* Participants that were below 18 or over 69, were dyslexic, did not speak Hindi, *or* were able to read Chinese or Japanese were automatically excluded from participation. During the exposure phase, the participants paid attention to the stimuli presented consecutively. Its completion took approximately 7 minutes. After the exposure phase, participants would see 7 real characters that showed up during the exposure phase (e.g. 土) as well as 7 pseudo-characters with the targeted phonetic radical (e.g. 扗) in randomized order, and were asked to judge what their pronunciations were likely to be. Since it would be quite challenging for non-Chinese speakers to explicitly pronounce their responses, the responses were instead elicited in the form of forced choice, among which the participants chose a pronunciation that they deemed fit. There were five options in each forced choice, which matched the number of alternative pronunciations for 土-bearing characters that were available in the exposure stimuli (namely, *du, mu, she, tu*, and *zao*). Each option came in the form of a sound button that repeated one pronunciation when clicked. Moreover, the association between each pronunciation and each button was randomized for each participant. Forced choice could bring the advantage of tapping into the implicit learning of the participants, who need not be aware of the reason behind their decision. Below the sound buttons were radio buttons that matched the labels of the sound buttons. For each real/pseudo character, participants made their choice by selecting one of the five radio buttons.

The experiment had been approved by the Faculty Ethics Assessment Committee Humanities (FetC-H) of Utrecht University, with the reference number 23-113-02.

3.2.4 Results

As mentioned in 3.2.2, there were 22 participants in total. Among them, 10 were automatically assigned to the 3R4E condition, while the other 12 were assigned to the 4R3E condition. Participants' responses were gleaned from the raw data, resulting in a total number of 308 data points. For subsequent statistical analysis, both selected responses and character type (real character vs pseudo-character) were parametrized in binary format. The response was encoded as 1 if it matched the dominant pronunciation, namely *tu* in both experimental conditions, and encoded as 0 if otherwise, namely if it matched any of the non-dominant pronunciations *du*, *she, mu,* and *zao*, all of which were outnumbered by the dominant pronunciation in both experimental conditions. Regarding character type, real characters were encoded as 1s, while pseudo-characters were encoded as 0s.

Like in Experiment 1, the statistical analysis here was also run on R. First, it was tested whether there was any significant difference in the proportion of dominant pronunciations between both experimental conditions. Since both conditions were not equal in their number of participants, the absolute number of 1s could not be counted upon. Instead, the *percentages* of dominant pronunciations in both conditions were used for comparison. As shown in Figure 2, the percentages of dominant pronunciations were similar across both conditions. Indeed, the two-sample test for equality of proportions suggested no significant difference ($p=0.7182$).

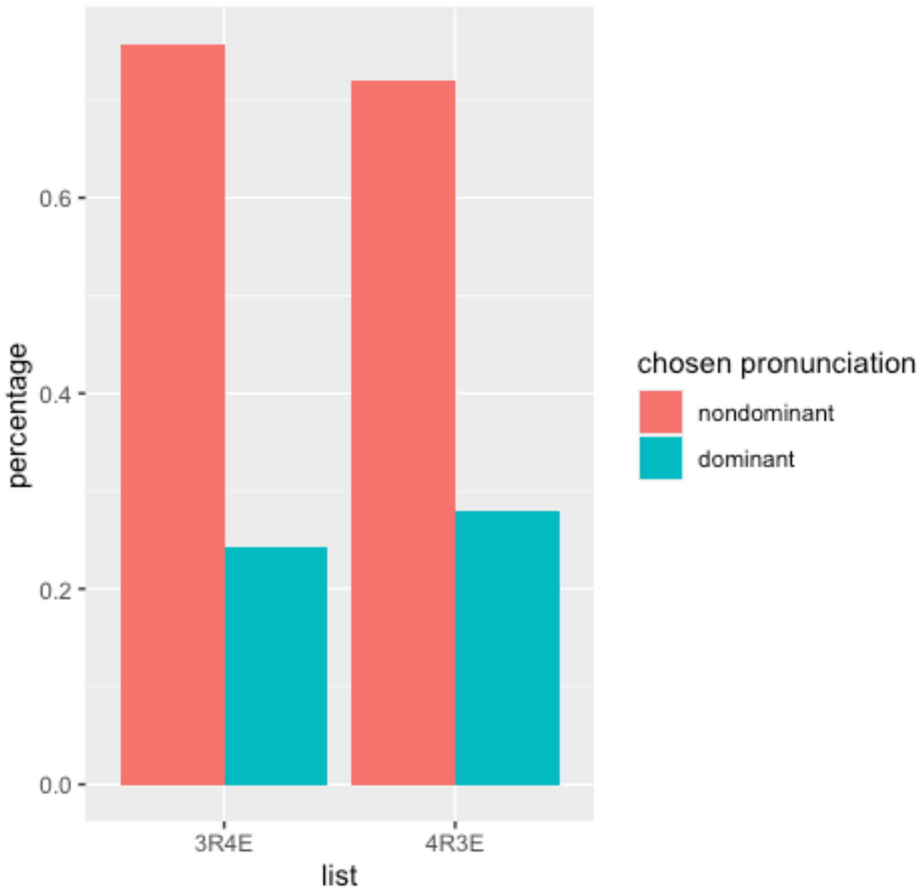

Figure 3: Percentages of Pronunciations in Both Experimental Conditions

Second, it was tested whether character type had any significant impact on the proportion of dominant pronunciations. Although participants without Chinese reading skills would be unable to tell apart real and pseudo-characters, it might be possible that prior exposure to real characters only would produce familiarity effect that biased participants' responses toward the more accurate recall or recognition of the pronunciations of real characters, and therefore less reliance on rule inference in comparison with out-of-the-blue pseudo-characters. However, as illustrated in Figure 3, the occurrence of dominant pronunciations was in fact *higher* for real characters than pseudo-characters[2]. Furthermore, the two-sample test for equality of proportions showed that there was no significant difference in the proportion of dominant pronunciations between real and pseudo-characters (p=0.1204). In other words, participants generally

[2] Given that there was an equal number of real and pseudo-characters in the response phase of the experiment, the results of both character types were directly comparable.

remained insensitive to the distinction between real and pseudo-characters despite repetitive exposure of the former.

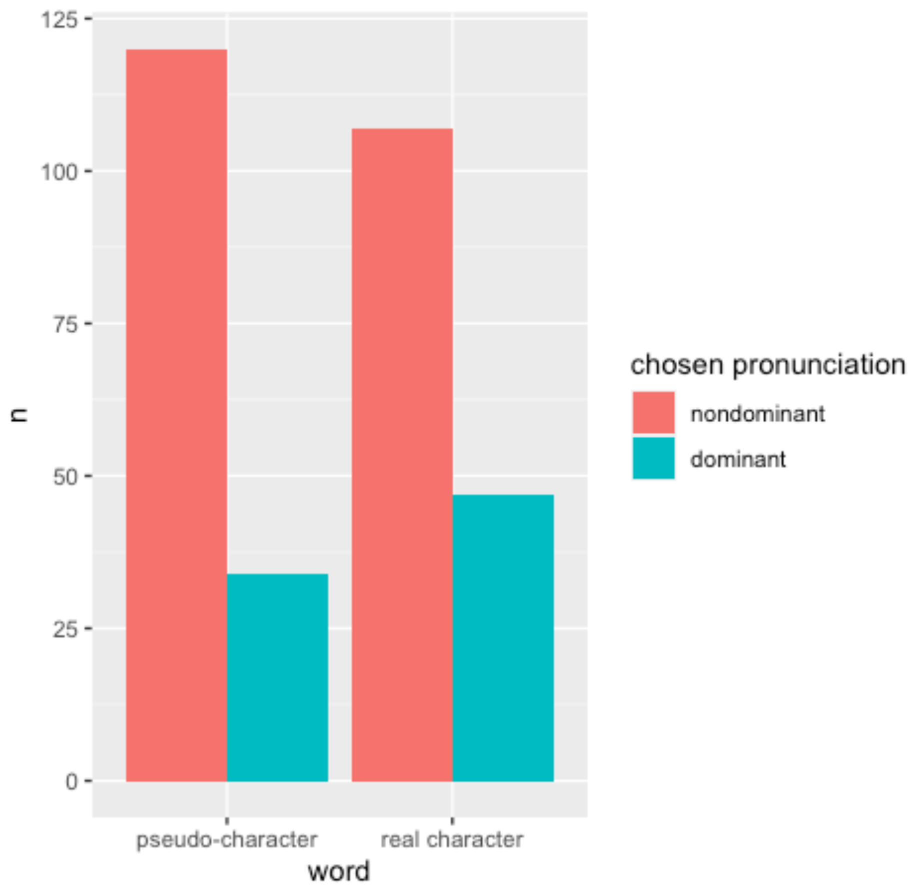

Figure 4: Percentages of Pronunciations Across Character Types

Next, it was tested whether the entire proportion of dominant pronunciations, regardless of experimental condition or character type, was significantly different from the product of random decision. If participants' decisions of the pronunciations of each character were random, then it was reasonable to expect that the probability of their choice of the dominant pronunciation (*tu*) would be approximately the same as other non-dominant pronunciations, namely 20% for each of the five options. In other words, here it was tested whether participants were likewise insensitive to the *dominance* of alternative pronunciations. It would be quite problematic if this turned out to be the case, because this would mean that the exposure phase, despite being replete with a preponderant amount of *tu*, had virtually no effect on participants' decision, signaling the failure of the exposure phase. Fortunately, the one-sample proportion test indicated that the ratio of pronunciations in this experiment, with a total of 81 dominant pronunciations and 227 non-dominant pronunciations, were indeed significantly different from 20% (p=0.007096). Thus participants were indeed influenced by the exposure phase, even though the effect did not vary between experimental conditions nor character types, as demonstrated by the previous proportion tests.

Finally, as in Experiment 1, several binomial generalized linear mixed models were built to analyze in a more sophisticated way whether experimental conditions or character types were able to explain data variance. The first mixed model, with participants and stimuli as random intercepts and experimental conditions as the fixed effect, was compared against the intercept-only model with only participants and stimuli as random intercepts. The ANOVA test suggested that the difference between both models was not significant (p=0.697). The second mixed model had both experimental conditions and character types as fixed effects, with the same set of random intercepts as the first mixed model. This model was then compared with the first mixed model. The ANOVA test suggested that these two models also did not differ significantly (p=0.1123). The third mixed model added random slopes to stimuli for both experimental conditions and character types. The ANOVA test showed that this model was not significantly different from the second model either (p=0.9997). The last mixed model included random slopes to participants for character types to the third model, since it was possible that different participants also had different sensitivity to the distinction between real and pseudo-characters. Still, the ANOVA test showed that it was not significantly different from the previous model (p=0.8876). Furthermore, the last two models did not converge, possibly due to similar technical issues encountered in the statistical analysis of Experiment 1 (see 3.1.4). Taken together, results consistently showed that neither experimental conditions nor character types exerted significant effect on participants' pronunciation decision.

4 Discussion

Results from Experiment 1 suggest that whether the associated dominant pronunciation of a phonetic radical is above or below its theoretical threshold as predicted by the Tolerance Principle makes little difference to its actual phonological assignment by Chinese readers. What are the possible explanations for this? One way to account for this is on methodological grounds. For example, the representativeness of the Gigaword2 corpus as the basis of this study may be questioned. The limited sample size of the experiment may also be insufficient as conclusive evidence. However, the more interesting theoretical implication would be that the Tolerance Principle does not seem to play a critical role in the formation of orthographic-phonological mapping rules. This does not directly refute the Tolerance Principle *per se*, but simply discourages its applicability to the reading domain. Reading, which recruits a complex of cognitive processes (Hillis & Caramazza 1992), may involve the synergistic work of multiple computational principles. In other words, the assumption that orthographic-phonological mapping is constrained by the Tolerance Principle may be overly simplistic. Alternatively, the Tolerance Principle could also be a strictly *linguistic* principle, without much bearing on other cognitive modules. Thus it is also possible that orthographic-phonological mapping crosses into territory beyond the scope of the Tolerance Principle.

On the other hand, results from Experiment 2 suggest that Hindustani-speaking participants' acquisition of Chinese orthographic-phonological mapping rules is likewise not constrained by the Tolerance Principle. There are multiple angles for the interpretation of the results. First, the fact that adults do not show categorical learning here confirms Schuler, Yang & Newport's (2021) observation that adults learn new linguistic rules in a probabilistic manner. It is worth pointing out that the adults in Experiment 2 *did* in fact acquire the preponderance of the dominant pronunciation (*tu*), as proven by statistical analysis, but nonetheless did not generalize its occurrence as part of a productive rule. Thus the explanation cannot simply be that adults do not learn from the experiment at all, but that the outcome of the acquisition does not match the predictions of the Tolerance Principle.

However, the fact that adults do not generalize the most frequent item across the board may also be due to an inherent bias in their decision-making when answering multiple choice questions, here provisionally termed as the *non-repetition bias.* Adults may have a tendency to *not* respond to all questions with the same answer, assuming that the experiment must not be so easy, or that each option must be chosen at least once for the diversification of the risk of non-accuracy. This bias, if exists, would be not unlike the “diversification bias” found in the literature of experimental psychology, according to which people tend to seek more variety in combined choices than separate choices (Read & Loewenstein 1995). The consecutive presentation of the multiple choice questions, or the experimental design as a whole, may have framed the questions as a form of combined choice, or a bundle of choices that are not entirely independent from each other. In other words, participants’ responses may crucially depend on “choice bracketing” (Read, Loewenstein, Rabin, Keren & Laibson 2000), specifically how participants perceived the relationship between choices. If combined choice is the way adult participants strategized during the experiment, then the attempt to elicit a clear-cut categorical response from them may become futile. Even if participants successfully grasp an orthographic-phonological rule, they may have *consciously* avoided producing a categorical response. In fact, this alternative explanation may further apply to Schuler, Yang & Newport’s (2021) findings, in particular the difference in performance between adults and children. It may be the case that, while adults were able to remain flexible with their responses, children may simply not have enough cognitive resources to reflect upon how each of their responses relates to each other, nor to produce responses that are more fine-grained and nuanced than a general rule (Anderson 2002, Dick 2014). Hence, instead of substantiating the exclusiveness of the Tolerance Principle to children, it is also possible that Schuler, Yang & Newport’s results simply reflect children’s limitation of their general executive function.

5 Limitations and Future Directions

This study is not without limitations. First, as noted in the previous section, this study raises potential methodological concerns, not only about the representativeness of the selected corpus and limited sample size, but also the issue with convergence in some of the mixed models. Thus further research with more corpora and data as well as more powerful statistics is needed to replicate and validate the findings of this study.

The use of traditional Chinese characters as experimental stimuli in this study also undercuts its generalizability. It is reasonable to expect that simplified Chinese characters, which show combinations of semantic and phonetic radicals different from traditional Chinese characters, would alter the parameters relevant for the Tolerance Principle (e.g. the type frequency of a phonetic radical and its associated pronunciations), and thus invite different predictions for each phonetic radical. For example, the traditional characters 漢 (*hàn* "Han Chinese") and 僅 (*jǐn* "merely") correspond to the simplified characters 汉 and 仅, but it is only in the simplified version that both characters share a phonetic radical, and that the new phonetic radical 又 amalgamates the type frequencies of both characters that would have been separate in traditional Chinese. Furthermore, the difference between traditional and simplified Chinese may also be complicated by the different general usage of words and collocational dependencies (e.g. "potato" is called 馬鈴薯 *mǎlíngshǔ* in Taiwan, but 土豆 *tǔdòu* in Mainland China). Thus it remains to be tested whether the Tolerance Principle is confirmed or disconfirmed in the case of simplified Chinese characters. More generally, the investigation of the possible role of the Tolerance Principle in reading should not be limited to logographic writing systems only. The acquisition of other varieties of writing systems, such as syllabary, abjad, alphabetic, abugida, and featural, could also be subject to the constraint of the Tolerance Principle. In any case, the relationship between the Tolerance Principle and writing must be looked at from a cross-linguistic perspective, and tested against a diversity of writing systems and the implications that follow.

Finally, as mentioned in the previous section, alternative explanations for the experimental results, such as the possible interaction between the Tolerance Principle and other linguistic principles that may obscure the clear-cut manifestation of the Tolerance Principle, or the conceivability that non-Chinese reading adults' performance in this study, as well as the behavioral distinction between children and adults in Schuler, Yang & Newport's (2021) experimental paradigm, may be an artifact of the methodology itself, remain to be addressed by future studies.

6 Conclusion

In sum, this study tests the applicability of Yang's Tolerance Principle to the reading domain, more specifically the reading of Chinese phono-semantic compounds. Experimental results of this study show that the Tolerance Principle does not constrain the orthographic-phonological mappings of phonetic radicals in Chinese readers, nor constrain its acquisition by non-Chinese readers. Inevitably, more research is needed, not only to extend or critique the findings of this study, but also to fine-tune methodological approaches for psycholinguistic studies, and, more generally, to advance the theoretical understanding of how language works in cognition.

7 Acknowledgment

For this study, I owe much gratitude to my supervisor Iris Mulders, who has given much feedback and instruction since the beginning of the study, in many informative discussions about theory and practice, that upholds the academic and methodological rigor of the study; Hugo Quené for his willingness to be the second reader for my thesis; Ben Bonfil for his strong technical support for the online experimental design, especially his meticulous debugging necessary for the improvement of both experiments; Kirsten Schutter for her helpful advice on statistical analysis, which clarifies the appropriate interpretation of the results; Rishabh Suresh for the recruitment of Indian participants via his social network, kindly doing so without recompense; and finally Tsai Tzu (Vanesa), not only for her substantial help with the recruitment of participants, but also for her considerate encouragement that constantly guides me through setbacks and frustration.

References

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child neuropsychology*, *8*(2), 71-82.

Audacity Team (2014). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 2.0.0 retrieved April 20th 2014 from http://audacity.sourceforge.net/.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry* 10: 533–581.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*(2-3), 150-177.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425-455.

Chen, M. J., & Weekes, B. S. (2004). Effects of semantic radicals on Chinese character categorization and character decision. *Chinese Journal of Psychology*.

Daniels, P. T. (2017). Writing systems. *The handbook of linguistics*, 75-94.

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351, https://joss.theoj.org/papers/10.21105/joss.05351.

Dick, A. S. (2014). The development of cognitive flexibility beyond the preschool period: An investigation using a modified Flexible Item Selection Task. *Journal of experimental child psychology*, *125*, 13-34.

Feldman, L. B., & Siok, W. W. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language*, *40*(4), 559-576.

Hernandez, A., Floyd, S., & Goldberg, A. E. (2001). Productivity depends on communicative intention and accessibility, not thresholds. *Threshold*, *1*, 1-1.

Hillis, A. E., & Caramazza, A. (1992). The reading process and its disorders. *Cognitive neuropsychology in clinical practice*, 229-261.

Ho, C. S. H., Ng, T. T., & Ng, W. K. (2003). A “radical” approach to reading development in Chinese: The role of semantic radicals and phonetic radicals. *Journal of literacy research*, *35*(3), 849-878.

Liu, P. D., Chung, K. K., McBride-Chang, C., & Tong, X. (2010). Holistic versus analytic processing: Evidence for a different approach to processing of Chinese at the word and character levels in Chinese children. *Journal of Experimental Child Psychology*, *107*(4), 466-478.

Liu, X., Vermeylen, L., Wisniewski, D., & Brysbaert, M. (2020). The contribution of phonological information to visual word recognition: Evidence from Chinese phonetic radicals. *Cortex*, *133*, 48-64.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, i–178.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 24.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Read, D., & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, *1*(1), 34.

Read, D., Loewenstein, G., Rabin, M., Keren, G., & Laibson, D. (2000). Choice bracketing. *Elicitation of preferences*, 171-202.

Schuler, K., Yang, C., & Newport, E. (2021). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient. https://doi.org/10.31234/osf.io/utgds.

Sketch Engine. (2017). Chinese Gigaword2 Corpus: Taiwan, Traditional [Data set]. https://www.sketchengine.eu/.

Sproat, R., & Gutkin, A. (2021). The taxonomy of writing systems: How to measure how logographic a system is. *Computational Linguistics*, *47*(3), 477-528.

Sturgeon, D. (2006). *Shuowen Jiezi* [Discussing writing and explaining characters]. Chinese Text Project. https://ctext.org/shuo-wen-jie-zi/zh.

Tsang, Y. K., Wu, Y., Ng, H. T. Y., & Chen, H. C. (2017). Semantic activation of phonetic radicals in Chinese. *Language, Cognition and Neuroscience*, *32*(5), 618-636.

Wittenberg, E., & Jackendoff, R. (2018). Formalist modeling and psychological reality. *Linguistic Approaches to Bilingualism*, *8*(6), 787-791.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.

Yang, C., & Montrul, S. (2017). Learning datives: The Tolerance Principle in monolingual and bilingual acquisition. *Second Language Research*, *33*(1), 119-144.

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

Zhou, X., & Marslen-Wilson, W. (2002). Semantic processing of phonetic radicals in reading Chinese characters. *Acta Psychologica Sinica*.

Zhou, L., Peng, G., Zheng, H. Y., Su, I. F., & Wang, W. S. (2013). Sub-lexical phonological and semantic processing of semantic radicals: a primed naming study. *Reading and Writing*, *26*, 967-989.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* Addison-Wesley, Cambridge, MA.