

Partial Replication of Grüter(2019): How Classifiers Facilitate Predictive Processing in L1 and L2 Chinese Speakers

How human utilize various cues from sources of information has been one of the central areas to understand human language processing. In this area, there is a claim that “L2 learners rely more on non-grammatical than on grammatical information during sentence processing compared to native speakers.(Grüter(2019))¹” To test this claim, Professor Grüter at the department of second language studies in the University of Hawaii finds an ideal language feature -- the sortal classifiers in Mandarin Chinese to conduct an visual world eye-tracking experiment. The advantage of Mandarin sortal classifiers is that they provide two concurrently available cues, grammatical form class and semantic information, so the researcher could find whether the listener allocates more weight to grammatical information when faced with two information sources. While the classifiers capture most nouns within the semantic information, some exceptions exist. For instance, the Chinese classifier 条Tiao is grammatically required when there is a numeral before the noun, and the nouns it classifies share similar characteristics like linear, long physical shape for Tiao. Fish and rope are both nouns that use the classifier Tiao, while wristwatch is also linear and long but does not use the classifier Tiao.

Provided with this phenomenon, Grüter et al selects three classifiers Zhi, Zhang, Tiao. For each classifier, four target items' pictures were chosen under the criteria of grammatically and semantically matched with the classifier. Grüter et al then designed three conditions' sets for each target item. Every set has three items presented: the target, the competitor, and a distractor. The condition is determined by the competitor's relationship to the classifier of the target. In the Tiao example, fish is one of the target items. It has a competitor basketball that is neither semantically nor grammatically aligned with the classifier Tiao as one condition. There is another condition of semantically aligned and grammatically not-aligned item toilet paper roll. The third condition is a semantically and grammatically acceptable object snake. With 36 critical trials in total, the researcher added 24 filler trials during the study. Participants were asked to listen to a short piece of audio and then choose one picture according to the information. Their eye movements were recorded to study their mean looking proportion to the

¹ Theresa Grüter, Elaine Lau & Wenyi Ling (2020) How classifiers facilitate predictive processing in L1 and L2 Chinese: the role of semantic and grammatical cues, *Language, Cognition and Neuroscience*, 35:2, 221-234, DOI: [10.1080/23273798.2019.1648840](https://doi.org/10.1080/23273798.2019.1648840)

target item area and competitor item area.

Participants were recruited in two groups: native Mandarin and non-native Mandarin speakers who have studied Mandarin. They are required to take a language proficiency test before starting the experiment. The number of participants recruited is 96, while the researchers exclude the data from 14 of them due to native languages (Cantonese) and poor eye gaze data. With data collected from eye-tracking, Grütter et al analysis shows that the non-native speaker have higher differences of fixation rate between on the G-S+ competitor (that's semantically aligned but grammatically not-aligned than on the target item than the differences shown on the native speakers). It validates the claim that non-native speaker of Chinese allocate more weights on semantic information than grammatical information during language processing tasks.

My partial replication of Grütter(2019) focuses on replicating the visual eye-tracking experiment for only Mandarin native speakers, as the Grütter(2019) data also exhibits that for native speaker, fixation pattern on G-S+ competitors is statistically different from pattern on G+S+ and G-S- (which are similar with the relative differences of percent fixation on target and competitors). I proposed a web-based simplification of the experiment with changes on theoretical emphasis on research topics, adjustment of visual stimuli representation, and a different method of gathering eye-tracking data online.

While the experiment is not involved with non-native Chinese speakers, it would not indicate any conclusions on second-language learner's language processing. Instead, it tries to replicate the result of native Chinese speakers' preference on grammatical cues. If this result is replicable, we should see a relatively similar pattern of fixation under condition G-S-(neither semantically nor grammatically matched with the classifier) and G-S+(only semantically aligned with the classifier of the target).

To run this experiment on web, I contacted Professor Grütter via email and then acquired all original visual stimuli and audio files of the experiment. I have also received support from Professor Degen and Leyla's template of an online eye-tracking experiment template. To fit with the template that exhibits two pictures for a page window, I cropped the original visual stimuli which was a picture with three items in one set to three separate pictures. I decided not to use a distractor pictures that are the same for every target item's set. In the online experiment, the participants would see two images with one on left and the other on right. They would hear the audio that is usually 2 to 3 seconds long to ask them to choose the target item pronounced. All

audio files remain unchanged. As the classifier and the fixation places after the onset of the classifier the key, Professor Grüter sets the audio distance between the onset of the classifier and the onset of the noun to be the same for every critical trial.

To start the experiment, the participants for my experiment would be asked to complete an eye-tracking accuracy test, for which they would click on red dots on the screen until they turn yellow with cameras on. Only when the participants have a score of accuracy higher than 50%, could they continue the tasks. They would not be able to randomly click on a trial until the audio has finished, which better presents the situation of people not doing the experiment in a serious manner. A language proficiency test is not included in the experiment, so it would be more important to look into the filler trials' data to check whether the accuracy score for fillers reflects whether the participants appropriately complete the experiment tasks. If using Prolific to post this experiment, there would be a first-language filter to select participants who are native speakers of Mandarin. As the note from Prolific said, there were about 38 participants who are satisfactorily qualified for my standard being active on the platform in past 90 days. The number of native speakers in this experiment is likely to be smaller than that in Grüter's original experiment.

During the process of preparing stimuli and setting up the experiment, there are some concerns of experiment design and execution I noticed. One classifier Zhi in Mandarin has three different forms of Chinese character (‘支, 只, 枝’) for classifiers. Each of them shares different meaning and noun categories. For instance, 支 Zhi usually describes the semantic categories that include pencils and candles with a slender cylinder physical form, while 只 Zhi refers to a different and wider variety of objects ranging from mice to balls, but they both share the same pronunciation including the tone. A confounding example in the experiment is the target item phone with the classifier Zhi as the aforementioned 支 Zhi and 只 Zhi are both grammatically matched with the object, which the pronunciation does not give an explicit choice of which classifier is being used until the participant hears the noun later.

Another issue is related to the experiment setup of image representation. With each target item, there are three different competitors for three sets. And the participants are supposed to choose the correct target images three times. While the experiment is capable to be set in a random mode to display all critical trials, it is likely that the participants run into the condition where they have consecutively listened to the same audio file and chosen the same target item

twice, which may increase the tendency to look at the same direction or the target object when a third trial begins.

The final concern is the format of filler trials. The visual stimuli are of the same style, but the audio played in fillers has differently formatted questions from critical trials. For example, critical trials only have the noun for the noun phrase after the classifier, while filler trials sometimes have adjectives and noun for the noun phrase. More importantly, there is no filler trials organized in the similar set as the critical trials. No repeated images were shown in filler trials, while critical trials have 12 items each repeated for 3 times. Some participants may realize or be influenced by recognizing this different pattern and make additional judgments that affect their process of language processing.