

Plausibility and Syntactic Reanalysis in Processing Novel Noun-noun Combinations During Chinese Reading: Evidence From Native and Non-native Speakers

Panpan Yao^{a*}, Reem Alkhamash^b and Xingshan Li^c

^aBeijing Language and Culture University, School of Psychology, Beijing, China

yaopp@blcu.edu.cn

^bEnglish Department, University College, Taraba, Taif University, P.O. BOX 11099, Taif 21944, Saudi Arabia

^cCAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China

Author note

This research was supported by Science Foundation of Beijing Language and Culture University (supported by “the Fundamental Research Funds for the Central Universities”) (21YBB33) and a grant from AThEME Project (Advancing the European Multilingual Experience, EU FP7 Grant 613465) to Panpan Yao, and Taif University Researcher Supporting Project (TURSP-2020/173) to Reem Alkhamash. The original data and codes are available from DOI 10.17605/OSF.IO/RPFYV.

Plausibility and Syntactic Reanalysis in Processing Novel Noun-Noun Combinations During Chinese Reading: Evidence From Native and Non-Native Speakers

Purpose

We aimed to tackle the question about the time course of plausibility effect in on-line processing of Chinese nouns in temporarily ambiguous structures, and whether L2ers can immediately use the plausibility information generated from classifier-noun associations in analyzing ambiguous structures.

Methods

Two eye-tracking experiments were conducted to explore how native Chinese speakers (Experiment 1) and high-proficiency Dutch-Chinese learners (Experiment 2) on-line process 4-character novel noun-noun combinations in Chinese. In each pair of nominal phrases (Numeral+Classifier+Noun1+Noun2), the plausibility of Classifier-Noun1 varied (plausible vs. implausible) while the whole nominal phrases were always plausible.

Results

Results showed that the plausibility of Classifier-Noun1 associations had an immediate effect on Noun1, and a reversed effect on Noun2 for both groups of participants.

Conclusions

These findings indicated that plausibility plays an immediate role in incremental semantic integration during on-line processing of Chinese. Similar to native Chinese speakers, high-proficiency L2ers can also use the plausibility information of classifier-noun associations in syntactic reanalysis.

Keywords: Chinese reading; classifier-noun association; eye tracking; plausibility effect; second language processing

Plausibility and Syntactic Reanalysis in Processing Novel Noun-Noun Combinations During Chinese Reading: Evidence From Native and Non-Native Speakers

Sentence reading is incremental, proceeding on a more or less word-by-word basis (Rayner et al., 2006). Some structures are inherently ambiguous in incremental semantic integration or syntactic parsing. A typical temporarily ambiguous structure that would cause a reanalysis process is subordinate-clause ambiguity, as illustrated in (1).

(1) As the woman edited **the magazine** about fishing amused all the reporters.

In processing such sentences, readers first interpret the main-clause subject (*the magazine*) as the object of the preceding verb (*edited*), only to revise this analysis when they encounter the disambiguating verb (*amused*) in the main clause (Adams et al., 1998; Ferreira & Henderson, 1991, 1993; Frazier & Rayner, 1982; Pickering & Traxler, 1998; Slatterly et al., 2013; Staub, 2007; Sturt et al., 1999). Pickering and Traxler (1998) found that this reanalysis process is influenced by the plausibility of the ambiguous noun phrase as the object of the preceding verb. When the verb was *sailed* instead of *edited*, which is implausible with *the magazine* being an object, readers spent a longer time in the ambiguous region (*the magazine*), but a shorter time in the disambiguating region (*amused*). This finding indicates that plausibility information plays an immediate role in incremental semantic integration, and it is easier to recover from the initial analysis when it is implausible than plausible

(Frazier & Clifton, 1998).

However, contradictory findings have been reported regarding whether singular nouns in temporarily ambiguous structures have a delayed or immediate plausibility effect. Using a word-by-word self-paced reading paradigm, Kennison (2005) explored the time course of the plausibility effect in sentence processing. Kennison pointed out that when encountering a singular noun following an adjective (e.g., *the ancient castle...*), the noun may or may not be the head noun that is modified by the adjective (e.g., *the ancient castle was...*), and the upcoming word may or may not be another noun that is a part of a noun compound (e.g., *the ancient castle gate...*). By manipulating the plausibility between the adjective and the following noun (*ancient castle* vs. *careful castle*), Kennison investigated whether the plausibility of singular nouns has an immediate effect in this temporarily ambiguous structure. The findings showed that the plausibility between an adjective and the following noun did not have an immediate influence on the noun. Specifically, reading times on the target nouns (*castle*) did not change significantly between the plausible (*ancient castle*) and implausible (*careful castle*) conditions. It seems that readers were waiting for more information to decide the grammatical role of the target noun *castle* instead of directly treating it as a head noun that is modified by the adjective *ancient*. Based on this finding, Kennison (2005) argued that singular nouns' plausibility information did not have an immediate role in temporarily ambiguous structures, and this was caused by the possibility of the singular noun being a modifier in noun-noun compounds. In other words, the plausibility effect is delayed in on-line processing of temporarily

ambiguous structures.

In contrast, Staub et al. (2007) found an immediate plausibility effect of singular nouns in temporarily ambiguous structures using the eye-tracking technique. They focused on readers' processing of noun-noun compounds in sentences such as (2a) and (2b).

(2a) The new principal talked to the **cafeteria manager** at the end of ...

(2b) The new principal visited the **cafeteria manager** at the end of ...

The compound *cafeteria manager* as a whole is fully plausible in both sentences, while the plausibility of the initial noun as a head noun varied with different preceding verbs (*talked to* vs. *visited*). They found that reading times (indexed by first fixation duration, gaze duration, and go-past time) on the initial noun (*cafeteria*) were significantly longer when the head noun analysis of this word was implausible (*talked to the cafeteria*) than when it was plausible (*visited the cafeteria*). The reanalysis procedure was triggered by the second noun (*manager*) which indicated that the initial analysis was incorrect. Staub et al. (2007) argued that these findings indicated an immediate plausibility effect on singular nouns in temporarily ambiguous structure processing, and the divergent findings between theirs and Kennison's (2005) were due to different experimental paradigms: compared to self-paced reading, eye-tracking technique can reflect participants' patterns in natural reading environments.

Unlike English, which is an alphabetic language, Chinese is a logographic script in which the text is formed by characters. There are no singular or plural markers on

nouns in Chinese or inflectional markers. The numeral meanings and grammatical categories of a certain word are highly dependent on the context (Chen, 1996; Chen et al., 2003). In addition, there are no explicit word-boundary markers in Chinese. Thus, Chinese readers need to rely largely on the context to decide whether a character is an independent word or a constituent morpheme of a multiple-character word (Chen, 1996, 1999). These script-unique features of Chinese are expected to cause more temporarily ambiguous structures and reanalysis processes relative to alphabetic languages like English, since the grammatical properties and meanings of each incoming character/word are controversial and dependent on the context (Liu et al., 2013).

Existing studies focusing on the plausibility effect of nouns in the on-line processing of Chinese multiple character strings reported divergent findings compared to English (Yang et al., 2012; Zhou & Li, 2021). Following Staub et al. (2007), Yang et al. (2012) investigated the time course of the plausibility effect in Chinese reading, focusing on 1-character words and 2-character compound words (which are composed of two 1-character words). In an eye-tracking experiment, Yang et al. (2012) manipulated the plausibility between the main verb and the first character of the following compound word/the following 1-character word. There were four conditions: the plausible-plausible condition as in (3a), in which the first character 门 (*door*) of the 2-character compound 门卫 (*gate-keeper*) is plausible with the preceding verb 踢打 (*kick*); the plausible-implausible condition as in (3b), in which the first character 门 (*door*) of the 2-character compound 门卫 (*gate-keeper*) is

implausible with the preceding verb 哀求 (*entreat*); the plausible condition as in (3c), in which the 1-character word 门 (*door*) is plausible with the preceding verb 踢打 (*kick*); and the implausible condition as in (3d), in which the 1-character word 门 (*door*) is implausible with the preceding verb 哀求 (*entreat*).

- (3a) Plausible-plausible 围观的人看着他踢打门卫却无动于衷
People were inattentive when he kicked the **gate-keeper**.
- (3b) Plausible-implausible 围观的人看着他哀求门卫却无动于衷
People were inattentive when he entreated the **gate-keeper**.
- (3c) Plausible 围观的人看着他踢打门却无动于衷
People were inattentive when he kicked the **door**.
- (3d) Implausible 围观的人看着他哀求门却无动于衷
People were inattentive when he entreated the **door**.

Yang et al. (2012) found a null plausibility effect in the compound noun region, as well as the first character of the compound noun. However, they found a plausibility effect on the 1-character word between (3c) and (3d). Yang et al. argued that these findings indicate that in Chinese, plausibility information can be immediately integrated in incremental processing, as indexed by the immediate plausibility effect between (3c) and (3d). The null plausibility effect between (3a) and (3b) implies that in Chinese, semantic information plays a role at the word level instead of the embedded character level. Specifically, when readers encountered 门 (*door*) in (3a) and (3b), they did not treat it as an independent word which should be a direct object of the preceding verb. Instead, they treated 门 (*door*) and the following

character 阂 (*gate*) as a whole unit. Thus, it was 阂 (gate-keeper), rather than 阂 (door), that was integrated with prior context.

Zhou and Li (2021) investigated how Chinese readers segment incremental 3-character words such as 酒精灯 (*alcohol lamp*) using eye-tracking technique. These incremental words can be treated as noun-noun combinations which are composed of a 2-character word 酒精 (*alcohol*) and a 1-character word 灯 (*lamp*). In their Experiment 1, the plausibility of the embedded 2-character word with the preceding verb varied (plausible vs. implausible) while keeping the 3-character compound words plausible in both conditions, as illustrated in (4).

(4) 陈晓默默地点燃/清洗酒精灯以便再次进行实验。

Chen silently lighted/cleaned **the alcohol lamp** in order to do the experiment again.

Zhou and Li (2021) found a null effect of plausibility in the initial 2-character word region, as well as in the 3-character compound word region. Consistent with Yang et al. (2012), Zhou and Li (2021) argued that their findings imply a long word advantage in Chinese lexical recognition: if multiple character strings can form a meaningful word, the long word is the first to be activated, not the embedded word (Li & Pollatsek, 2020). In other words, the semantic information of embedded words is not activated, and thus has no effect on integration.

The consistently observed null plausibility effect of the initial constituent of noun-noun compounds in Chinese (Yang et al., 2012; Zhou & Li, 2021) is divergent from the findings in English (Staub et al., 2007). There are some possibilities for this

divergence. The first is the type of compound words. Compound words can be further divided into two groups: familiar compound words (which are known to readers) and novel compound words (which may be unfamiliar to readers; however, readers can immediately get their meanings by combining the meanings of the two components). Previous studies have found that familiar compound words are represented and processed as whole units (like lexicalized words that have corresponding independent entries in the mental lexicon), while novel compound words are compositionally stored and accessed (Hyönä, 2015; Hyönä et al., 2020). The compound words in Staub (2007) were novel compound words¹ that had not been lexicalized and needed to be accessed compositionally through each constituent, whereas the 2- and 3-character compound words in Yang et al. (2012) and Zhou and Li (2021) were familiar compound words that had been lexicalized and could be found as independent entries in a dictionary. According to Li and Pollatsek (2020), in Chinese reading, the constituents of familiar noun-noun compounds cannot be sufficiently activated to participate in integration. Thus, it is reasonable to observe an immediate plausibility of the initial constituent of compound words in Staub et al. (2007), but not in Yang et al. (2012) and Zhou and Li (2021). The second possibility is the word length of compounds/combinations between English and Chinese. Both 2- and 3-character words in Chinese are shorter than noun-noun compound words in English

¹ In Staub et al. (2007), they used novel compound words (e.g., *cafeteria manager*) and familiar compound words (e.g., *mountain lion*), although it is not clear whether or not their ‘familiar compound words’ have been lexicalized and have their own independent entries in a dictionary as defined in the current research. Nonetheless, these two groups of compound words did not differ significantly regarding the plausibility effect on the initial noun region in Staub et al. (2007), since they reported null compound type by plausibility interaction on any region.

(门卫/酒精灯 vs. *cafeteria manager*). According to the visual acuity principle proposed by Bertram and Hyönä (2003), compound words may be processed holistically when they are short enough to fit in the foveal area of the eyes where vision is most acute. As postulated by Yang et al. (2012), it is possible to observe a plausibility effect on the first 2-character constituent when extending the word length of noun-noun combinations to four characters. The third possibility is the different scripts between Chinese and English. In reading English, one can easily tell that there are two words in *cafeteria manager* because of the space between *cafeteria* and *manager*. When incrementally processing this compound, readers can take the space between the words as a cue to analyze them separately and compositionally. In Chinese, however, due to the lack of word boundary markers and inflectional markers, plausibility information may work in a different manner.

Considering the script-unique features of Chinese, there are two contrasting accounts regarding the time course of the plausibility effect in the on-line processing of Chinese. A delayed account argued that semantic contextual information would not be expected to function immediately (Aaronson & Ferres, 1986). In reading Chinese, a delayed comprehension strategy would be more suitable and efficient for building up a coherent integration because it would maximize the amount of available information and minimize the potential ambiguity and reanalysis given the uncertainty of an input's grammatical category and meanings. Conversely, an immediate account argued that, although with different scripts, semantic information can be used automatically and immediately in the on-line processing of Chinese sentences, just as

in alphabetic languages (Wang et al., 2008; Yang et al., 2009).

Focusing on cross-script comparison, the current research aimed to test the third possibility (script features between English and Chinese) while keeping the compound word type and length in a more analogous way to Staub et al. (2007). In Experiment 1, we used the eye-tracking technique to investigate whether there is an immediate plausibility effect when processing 4-character novel noun-noun combinations (which are composed of two 2-character words). Unlike Yang et al. (2007) and Zhou and Li (2021), in which all the compound words are lexicalized words and can be found in Chinese dictionaries as independent entries, the 4-character noun-noun compounds/combinations used in the current study are novel and do not have their own lexical entries. Li, Rayner, and Cave (2009) found that when processing a 4-character string composed of two semantically related 2-character words (in their Experiment 5), the first 2-character word would be processed earlier than the second one. Furthermore, according to Hyönä et al. (2020), novel compound words are decompositionally stored and accessed. Thus, the 4-character novel noun-noun combinations in the current study are expected to be processed decompositionally through two constituents. Each noun-noun combination was embedded in a nominal phrase with the structure [Numeral+Classifier+Noun1+Noun2], in which the plausibility of the initial constituent noun (Noun1) with the preceding classifier was manipulated to be either plausible or implausible. According to the delayed account that script-unique features of Chinese may require a delayed integration strategy to minimize potential ambiguity and reanalysis, a null plausibility effect on the initial

constituent would be expected. However, according to the immediate account that semantic information can be used automatically and immediately in the on-line processing of Chinese sentences, just as in alphabetic languages (Wang et al., 2008; Yang et al., 2009), an immediate plausibility effect on the initial constituent would be expected.

In addition to exploring how native Chinese speakers on-line process 4-character noun-noun combinations, we aimed to further investigate whether Dutch speakers who learned Chinese as a second language (henceforth, Dutch-Chinese learners) can use plausibility information of nouns in real-time processing in a native-like pattern, especially when this plausibility information is generated from a unique-to-Chinese structure ---- classifiers. Different from alphabetic languages, a classifier is obligatory every time when combining a numeral with a noun in Chinese. Due to the lack of classifiers in alphabetic languages such as Dutch, all the meanings and distributions of classifiers, as well as the associations between classifiers and proper nouns, are unique-to-L2 knowledge for Dutch-Chinese learners.

Theories differ regarding the ultimate achievement of late L2 acquisition. The Critical Period Hypothesis (Lenneberg, 1967) states that due to maturational reasons, late L2ers (L2ers who started to learn L2 after puberty) have difficulty in acquiring and processing some aspects of L2 (see a review by Slabakova, 2016). Furthermore, the Morphological Congruency Hypothesis (Jiang et al., 2011) argues that L2 acquisition and processing are influenced by morphological congruence between L2ers' L1 and L2. Only congruent L2 knowledge (the morphemes that exist in both

L1 and L2) can be fully acquired by high-proficiency late L2ers. For incongruent L2 knowledge (the morphemes that exist only in L2 but not in L1), it is extremely difficult, if possible, to develop a native-like representation. On the other hand, some researchers argued that late L2ers' grammatical representations and language processing routines are quantitatively but not qualitatively different from native speakers. They can eventually acquire native-like L2 representations with increased proficiency levels (Hopp, 2006; White, 2003). In addition, the Feature Reassembly Hypothesis (Lardiere, 2009) proposes that the restriction from L1 is limited and can be overcome with an increase in L2 proficiency. In other words, some incongruent and unique-to-L2 constructions can be ultimately acquired by late L2ers through the "feature reassembly" procedure.

Previous studies offered inconsistent findings regarding the ultimate acquisition of unique-to-L2 constructions. Some studies found that late L2ers exhibit non-native-like behavior patterns with unique-to-L2 structures (Chen et al., 2007; Ionin et al., 2012; Jiang, 2004, 2007; Kotz, 2009). In contrast, native-like behaviors in late L2ers have been found with unique-to-L2 constructions in behavioral experiments (Hopp, 2006; Ionin & Montrul, 2010; Jackson, 2007; Jackson & Dussias, 2009), eye-tracking studies (Yao & Chen, 2016), and ERP studies (Frenck-Mestre et al., 2009; Morgan-Short et al., 2010). So far, it is still unclear based on these existing findings whether late L2ers can eventually exhibit native-like behaviors in the processing of unique-to-L2 constructions. Considering the uniqueness of the classifier system in Chinese compared to alphabetic languages such as Dutch, investigating whether

high-proficiency late Dutch-Chinese learners can use the plausibility of classifier-noun associations in structure reanalysis could shed some light on the debate on the ultimate L2 acquisition of unique-to-L2 constructions (the Morphological Congruency Hypothesis vs. the Feature Reassembly Hypothesis). We tackled this question in Experiment 2, in which the same materials and procedures as in Experiment 1 were used for high-proficiency late Dutch-Chinese learners.

Previous studies found that in processing temporarily ambiguous structures, L2ers behaved like native speakers with an initial analysis that could be corrected and reanalyzed later (Juffs & Harrington, 1996). Compared to native speakers, L2ers find it more difficult to recover from initial syntactic misanalysis (Hopp, 2015; Roberts & Felser, 2011; Williams et al., 2001). In addition, both native and non-native readers are sensitive to plausibility information during processing (Williams et al., 2001; Williams, 2006). Based on these previous studies, we predict that if late L2ers can eventually acquire unique-to-L2 knowledge (as proposed by the Feature Reassembly Hypothesis), native-like eye movement patterns are expected for high-proficiency Dutch-Chinese learners. In other words, L2ers are expected to use the plausibility information of classifier-noun associations immediately in on-line processing. However, if late L2ers cannot acquire unique-to-L2 knowledge (as proposed by the Morphological Congruency Hypothesis), they would exhibit different patterns compared with native Chinese speakers.

Experiment 1

In this experiment, we investigated whether there is an immediate plausibility effect when incrementally processing 4-character novel noun-noun combinations in Chinese.

Participants

Sixty college students (30 female, age range 18-35 years, $M_{age} = 22.70$, $SD_{age} = 3.95$) in China participated in the experiment². All participants were native speakers of Mandarin Chinese and had normal or corrected-to-normal vision.

Materials

There were 30 pairs of nominal phrases with the structure [Numeral+Classifier+Noun1+Noun2], in which Noun2 was always the head noun and Noun1 was always the modifier of Noun2. Each pair of nominal phrases varied only in Noun1. None of these 60 novel noun-noun combinations (Noun1+Noun2) can be found in any Chinese dictionary or lexical corpus as an independent entry, indicating that they were not lexicalized. The plausibility between the classifier (Cl) and Noun1 varied in two conditions: plausible and implausible³, while the nominal phrase as a whole (as well as the sentences they were embedded in) was always plausible⁴. For

² The number of participants was decided based on the results of the power analysis on a pilot study of 40 native Chinese speakers with the exact same materials and procedure. These 40 participants' gaze duration on Noun1 were analyzed using linear mixed effect model focusing on the effect of plausibility condition. The powerSim function (Judd et al., 2012) was used to test the statistical power of this model. Results showed that the power would reach .85 with 60 participants.

³ The plausibility and acceptability of the Cl-Noun1 associations was normed by 10 native Chinese speakers using a 5-point scale, in which 1 stands for implausible while 5 stands for plausible. The averaged norming score of the Cl-Noun1 associations in the plausible condition was 4.85, which is significantly higher than that in the implausible condition ($M = 1.12$), $p < .001$.

⁴ The plausibility and acceptability of the nominal phrases in general was normed in a similar

example, although the nominal phrases 一件毛衣外套 (*one piece of sweater coat*) in (5a) and 一件羊毛外套 (*one piece of woolen coat*) in (5b) are both grammatical and acceptable in general, Noun1 毛衣 (*sweater*) in (5a) is plausible with the Cl 件 (*piece*), while Noun1 羊毛 (*wool*) in (5b) is implausible with the same Cl. As illustrated in Table 1, the descriptive features (e.g., frequency and number of strokes) of Noun1 in the two conditions were matched. In addition, the bigram frequency of Noun1 and Noun2 were matched between the two conditions. In total, 60 critical sentences were pseudo-randomly divided into two lists to ensure that each list contained only one condition (either plausible or implausible) for each pair. In addition, there were 120 fillers and five practice trials in each list. None of the filler sentences contained nominal phrases with the structure [Numeral+Classifier+Noun].

(5a) Plausible 柜子 里的那 一件 **毛衣** 外套 是妈妈 新买的。
 Wardrobe inside that one Cl_piece **sweater** coat is Mom new buy
 That sweater coat in the wardrobe was newly bought by Mom.

(5b) Implausible 柜子 里的那 一件 **羊毛** 外套 是妈妈 新买的。
 Wardrobe inside that one Cl_piece **wool** coat is Mom new buy
 That woolen coat in the wardrobe was newly bought by Mom.

[Insert Table 1 here]

way by 10 native Chinese speakers. The results indicated that nominal phrases in two conditions were equally plausible and acceptable to native Chinese speakers: $M_{plausible} = 4.61$, $M_{implausible} = 4.67$, $p = .54$. Furthermore, the familiarity and acceptability of the Noun1+Noun2 combinations was normed by a different group of 10 native Chinese speakers using a 5-point scale, in which 1 stands for very strange and unacceptable, while 5 stands for familiar and acceptable. The results indicated that these combinations in two conditions did not differ significantly in their familiarity and acceptance: $M_{plausible} = 4.65$, $M_{implausible} = 4.74$, $p = .28$.

Apparatus

Participants' eye movements were recorded using an Eyelink 1000 eye tracker with a sampling rate of 1000 Hz. Each sentence was presented in one line in the middle of a 21-inch cathode ray tube (CRT) monitor. Participants were seated 65 cm away from the monitor. Following a 3-point horizontal calibration and validation, the gaze-position error was less than 0.5°. Chinese Song font was used for all materials. Each character occupied a 36 × 36 pixel grid with one character subtending approximately 1° of visual angle. Eye movements were recorded from the right eye, but the viewing was binocular.

Procedure

Participants were tested individually. The experiment started with a brief instruction and a standard horizontal three-point grid calibration and validation. Then, five practice trials were conducted to ensure that the participants understood the task and were familiar with the apparatus. Critical experimental trials were conducted after practice trials. Participants read sentences silently on the screen. At the beginning of each trial, a drift check was conducted. Each sentence appeared after participants fixated on a character-sized box at the location of the first character of each sentence. After reading each critical sentence, participants were asked to press a button so that the original sentence disappeared and was replaced by a meaning-related question, to which participants responded by pressing a button. During the eye-tracking experiment, a calibration was conducted every time participants had problems performing the drift check. The entire experiment lasted 30 minutes.

Results and Discussion

The nominal phrases were divided into three regions: the Num+Cl region, which includes the number — (*one*) and the classifier; the Noun1 region; and the Noun2 region. Seven eye movement indices were analyzed for each region. *First fixation duration (FFD)* is the duration of the first fixation in a region. *Gaze duration (GD)* is the sum of the fixation durations before the eyes first move out of a region. *Go-past time (GP)* is the duration from first entering a region to first going past it forwardly. *Total reading time (TT)* is the sum of the durations of all fixations in a region. *Skipping rate (SKR)* is the probability of a region being skipped. *Regression-in probability (RI)* is the proportion of trials where readers looked back from the right after they first passed a region. *Regression-out probability (RO)* is the proportion of trials where readers looked back to an earlier point after they fixated on this region.

The mean response accuracy was 98%. Fixations shorter than 80 ms or longer than 1000 ms were removed. Linear mixed effects (LME) models (Baayen et al., 2008) were constructed for FFD, GD, GP, and TT, and a generalized LME model was used for SKR, RO, and RI. The plausibility of the Cl-Noun1 association was treated as a fixed effect (with the factor levels coded as -.5 and .5 for plausible and implausible conditions, respectively), specifying participants and items as crossed random effects, including intercepts and slopes. The random effect structure of the models was trimmed down when the models failed to converge, in accordance with Bates et al. (2015). The statistical procedure was conducted using the *lmer* function (Bates et al., 2011) and *lmerTest* function (Kuznetsova et al., 2017) in *R* (version 3.5.1; R Core

Team 2018). The means and SEs for each eye movement index in all three regions are summarized in Table 2, and the results of the statistical analysis are summarized in Table 3⁵.

[Insert Table 2 & Table 3 here]

The Num+Cl region

We analyzed the effect of plausibility manipulation in this region to examine the parafoveal-on-foveal effect, which refers to the extent to which properties of the word to the right of fixation exert an influence on the currently fixated word (see Rayner, 2009, for discussion). In the Num+Cl region, no significant difference was found between the plausible and implausible conditions for any eye movement measure. This finding indicated that there was no parafoveal-on-foveal effect in this region.

The Noun1 region

In the Noun1 region, there was a significant plausibility effect on GD and SKR, with longer fixation durations and lower skipping rate on Noun1 in the implausible condition than in the plausible condition. These findings indicate that on reading Noun1, participants treated it as a head noun which should be associated with the preceding Cl. In this case, processing costs occurred in the implausible condition compared to the plausible condition. In on-line Chinese reading, the plausibility

⁵ The log-transformed data of fixation measures yield the same pattern of statistical significance as the analysis based on the raw data. In the interest of transparency of effect sizes, we report the analysis of the untransformed data here.

information of nouns plays an immediate role in semantic integration. In addition, significantly more regressions were directed towards Noun1 in the plausible condition than in the implausible condition.

The Noun2 region

In the Noun2 region, there was a reversed plausibility effect on GD, TT, and RO, with longer fixations and higher regression-out probability on Noun2 in the plausible condition than in the implausible condition. We postulate that these findings may indicate that on reading Noun2, participants realized that their initial analysis of treating Noun1 as a head noun was incorrect. They need to give up the initial analysis and reanalyze Noun2 as the head noun, and treat Noun1 as a modifier that modifies Noun2. When Noun1 was plausible with C1 (the plausible condition), it costs more to recover from the initial analysis compared to the implausible condition.

In summary, we found a plausibility effect on Noun1 and a reversed plausibility effect on Noun2. These findings indicate that the plausibility effect has an immediate role in the on-line processing of novel noun-noun combinations.

Experiment 2

In this experiment, we explored whether high-proficiency Dutch-Chinese learners can use the plausibility information of classifier-noun associations in processing 4-character noun-noun combinations in Chinese.

Participants

Thirty Dutch-Chinese learners⁶ (16 female, aged 22 to 35 years, $M_{age} = 25.53$, $SD_{age} = 3.53$) participated. They were graduate students from Leiden University. They all started learning Chinese after puberty. All of them had been learning Chinese for at least three years and had the experience of studying in China for at least six months ($M_{China\ living\ experience} = 2.47$ years, $SD = .75$). All L2 participants had passed the HSK-C1 (Hanyu Shuiping Kaoshi - advanced level, the standard Chinese language proficiency test for non-native speakers administered by the Ministry of Education of the People's Republic of China), indicating that they were high-proficiency L2-Chinese learners.

Materials

Same as in Experiment 1⁷.

Apparatus

Same as in Experiment 1.

Procedure

⁶ The number of L2 participants in Experiment 2 (30) was less than the number of native participants in Experiment 1 (60). This was due to the fact that there were limited number of available high-proficiency Dutch-Chinese learners. However, this did not mean that the data of L2 participants was under-powered. L2 participants' TT on Noun1 were analyzed using linear mixed effect model focusing on the effect of plausibility condition. The powerSim function (Judd et al., 2012) was used to test the statistical power of this model. Results showed that the power reached .80 with 30 L2 participants.

⁷ Two off-line tests were conducted before the critical eye-tracking experiment to make sure that all L2 participants were familiar with the critical nouns (both Noun1 and Noun2) and the Cl-Noun1 and Cl-Noun2 associations used in the current study. The first test was a lexical definition test in which participants were asked to explain the meanings of some Chinese nouns in both Dutch and Chinese. The second test was a gap-filling test in which participants were asked to fill in proper classifiers in the brackets in nominal phrases such as [一 () 毛衣] (*one () sweater*). Dutch-Chinese learners whose accuracy in these two tests were higher than 90% were selected as our L2 participants. Through these two tests, we were assured that all L2 participants in the current experiment have already acquired the basic and necessary lexical knowledge of the classifiers and nouns used in the current research.

Same as in Experiment 1.

Results and Discussion

Similar to Experiment 1, seven eye movement indexes (FFD, GD, GP, TT, SKR, RI, and RO) in the three regions were analyzed. The mean response accuracy of the 30 participants was 96%. The methods of data cleaning and modeling were the same as those in Experiment 1. The means and SEs for each eye movement index in all three regions are summarized in Table 2, and the results of the statistical analysis are summarized in Table 4.

[Insert Table 4 here]

The Num+Cl region

There was no significant difference between the plausible and implausible conditions for eye movement measures, except for RI. Compared to the implausible condition, more regressions were directed towards the Num+Cl region in the plausible condition.

The Noun1 region

In the Noun1 region, there was a significant plausibility effect on GD, TT, and SKR, with longer fixations and lower skipping rate on Noun1 in the implausible condition than in the plausible condition. These native-like patterns indicate that, similar to native speakers, L2ers used the plausibility information of Cl-Noun1 associations immediately in on-line processing.

The Noun2 region

In the Noun2 region, there was a reversed plausibility effect on GP, TT, and RO, with significantly longer fixations and higher regression-out probability on Noun2 in the plausible condition than in the implausible condition. Again, L2ers were patterned with native speakers. These findings indicate that, similar to native speakers, L2ers experienced a garden-path process when incrementally processing Chinese noun-noun combinations.

In summary, similar to native Chinese speakers, we found a plausibility effect on Noun1 and a reversed plausibility effect on Noun2 for high-proficiency Dutch-Chinese learners. These findings indicate that, just as native speakers, Dutch-Chinese learners process Chinese noun-noun combinations incrementally. Furthermore, they have acquired the unique-to-L2 knowledge and can use the plausibility information of classifier-noun associations in structure reanalysis.

General Discussion

Two eye-tracking experiments were conducted to explore how native Chinese speakers (Experiment 1) and high-proficiency Dutch-Chinese learners (Experiment 2) on-line process 4-character novel noun-noun combinations in Chinese reading. By doing so, we aimed to tackle the question about the time course of the plausibility effect in on-line processing of Chinese nouns, and about whether L2ers can use the plausibility information generated from a unique-to-L2 structure in ambiguous structure analysis.

In Experiment 1, we found a significant plausibility effect on Noun1 on early processing measures GD and SKR, which indicated that when first reaching Noun1, native speakers immediately interpreted it as a head noun that should be associated with the preceding CI, and integrated it with preceding fragments directly. Instead of waiting for further information to determine the meaning and grammatical role of Noun1, participants' immediate use of the plausibility information of CI-Noun1 associations supports the immediate account. In Noun2 region, we found a significant reversed plausibility effect on GP, TT, and RO. We argue that this was caused by the processing costs of the reanalysis procedure. Specifically, when reaching Noun2, participants realized that their initial analysis of treating Noun1 as a head noun was incorrect. They must give up the initial analysis and reanalyze the structure. Since participants were more strongly committed to the initial analysis in the plausible condition than in the implausible condition, they were less willing to abandon it in the face of new evidence (Frazier & Clifton, 1998; Pickering & Traxler, 1998). This is consistent with the pattern of RI in Noun1 region: more regressions were directed towards Noun1 in the plausible condition than in the implausible condition, indicating that it was more likely for participants to look back to Noun1 for more information to reanalyze the nominal phrase structure when encountering Noun2 in the plausible condition than in the implausible condition. This reversed plausibility effect observed in Noun2 also supports the immediate plausibility effect of Noun1.

The cognitive mechanism of this reversed plausibility effect on Noun2 is analogous to that of task-switching. Previous studies found a typical switch cost when

participants were required to accomplish switch trials (i.e., two successive trials that require different rules) compared to repeat trials (i.e., two successive trials applying the same rules): longer reaction times and lower accuracy on the former than the latter one (Crone et al. 2004; Gajewski et al., 2010). This indicates that individuals tend to carry over the previous rule/logic, thus causing processing failure when the previous rule/logic is no longer appropriate. In other words, inhibiting prepotent rules/logic requires a cost. The reversed plausibility effect on Noun2 in the current study could be treated as an indication of this processing failure. Specifically, due to the incremental and immediate semantic integration, in the plausible condition, the Cl-Noun1 connection was built before the occurrence of Noun2. On reaching Noun2, where the Cl-Noun1 connection needs to be broken, inhibiting and abandoning a plausible and strong Cl-Noun1 connection should be difficult and thus would cause some processing costs. However, in the implausible condition, there would be no legit Cl-Noun1 connection, and participants thus may expect the upcoming item (Noun2) to be an appropriate item that can be associated with Cl. In this case, on reaching Noun2, there would be no pre-built rule/logic that needs to be inhibited/abandoned, and the plausibility of the Cl-Noun2 connection would make the processing of Noun2 smooth and quick.

Previous studies found a null plausibility effect on Noun1 in Chinese (Yang et al., 2012; Zhou & Li, 2021), while we found a plausibility effect in the current study. This divergence can be explained by the following possibilities: The first is the type of compound/combinations. In the current research, we used novel noun-noun

combinations that are not lexicalized and need to be accessed compositionally. In Yang et al. (2012) and Zhou and Li (2021), however, the key materials were familiar compound words/incremental words that were lexicalized and accessed holistically. Thus, the divergent findings in the current study and previous ones may reflect the different processing mechanisms between compounds that are stored holistically and combinations that are compositionally stored and accessed. The second possibility is the word length of the compounds/combinations. According to previous studies (Inhoff & Liu, 1998; Yan et al., 2015), the perceptual span in Chinese varied from three to four characters on the right of the fixation. In Yang et al. (2012) and Zhou and Li (2021), the compounds have two to three characters that can fall in one perceptual span and thus be processed simultaneously. In the current study, the four characters of Noun1+Noun2 combinations may not fall within one perceptual span when readers fixate on the Num+Cl region. In other words, it may be more possible to access the meanings of two or three characters in parallel than four characters within a certain fixation. The current research cannot distinguish these two possibilities. Future research is required to address this question. For example, studies that compare novel compounds with familiar compounds while controlling word length may offer some insights into this issue⁸.

In addition, the third possible reason for the divergent findings is the syntactic context of material sentences. In Yang et al. (2012) and Zhou and Li (2021), the preceding context (the critical verb, similar to the materials in Staub et al., 2007)

⁸ We thank an anonymous reviewer for this suggestion.

supports general predictions about word class and thematic roles of the following nouns, which may be less constraining than the classifiers in the current research that restrict word class, thematic roles, animacy, and shape of the upcoming nouns. Thus, an anomalous classifier-noun association may evoke stronger reactions than an anomalous verb-noun combination⁹. Furthermore, the divergent findings may also be evoked simply by the rather low occurrences of 4-character words. According to Lexicon of Common Words in Contemporary Chinese Research Team (2008), 72% of Chinese words are 2-character words, 12% are 3-character words, and 10% are 4-character words. Moreover, novel 4-character combinations by nature would have an even lower occurrence than lexicalized 4-character words. Thus, it could be the case that readers' default processing works in a holistic fashion (when processing 2/3-character words), whereas their processing of novel 4-character words which representing a special case works in a decompositional fashion¹⁰. Future research is needed to distinguish these possibilities.

In this study, we found similar processing patterns on Noun1 of noun-noun combinations as those reported by Staub et al. (2007). This indicates that when controlling compound word type and word length, the incremental semantic integration is similar in Chinese and English, even with the lack of word boundary and inflectional markers in Chinese. In both languages, plausibility plays an immediate role in on-line processing. Interestingly, compared to Staub et al. (2007), the current research obtained some different findings on Noun2. No processing cost of

⁹ We thank an anonymous reviewer for this suggestion.

¹⁰ We thank an anonymous reviewer for this suggestion.

the reanalysis procedure was reported by Staub et al. (2007), whereas in the current study, we observed a reversed plausibility effect which we speculated reflects the reanalysis cost. We think this divergence may be caused by some possibilities. The first is the different processing mechanisms in different languages. When parsing Chinese, readers have a comparatively heavier reliance on context semantic information than parsing alphabetic languages (due to the lack of explicit word boundary and inflectional markers in Chinese). Specifically, compared to reading alphabetic languages, in reading Chinese, readers may commit more strongly to the plausible initial analysis, which would cost more when they have to abandon it and reanalyze the structure. Thus, even though the nominal phrases in general were plausible and grammatical in both conditions, processing difficulties were revealed on Noun2 in the plausible condition compared to the implausible condition in the current study. Furthermore, this divergence may also be caused by the syntactic context of material sentences. As discussed before, the preceding context in Staub et al. (2007) (i.e., the verb-noun associations) offered fewer constraints than that in the current study (i.e., the classifier-noun associations). Thus, recovering from a classifier-noun association may evoke stronger reactions and require more time and resources than recovering from a verb-noun combination. From this perspective, it is reasonable to observe a reversed plausibility effect on Noun2 in the current study but not in Staub et al. (2007). Future research is needed to explore these possibilities.

In general, the results of Experiment 1 indicated that the real-time processing of 4-character Chinese novel noun-noun combinations is incremental. More precisely,

the integration of each incoming character/word is serial and immediate. This finding offers some important implications for sentence-processing models. According to the E-Z Reader model, words are processed serially so that the processing of word n does not start until the processing of word $n-1$ has been accomplished. This serial processing method is implemented in both lexical identification and semantic integration (Reichle et al., 2009). However, this model was built based on alphabetic languages and did not consider the unique script feature of Chinese ---- the lack of word boundary marker. It is not very likely for Chinese readers to process one word at a time, since they would not know where word boundaries are before the words within the perceptual span have been identified. Considering this Chinese script feature, the Chinese Reading Model (CRM) argued that all words within the perceptual span are activated in parallel and compete with each other for identification (Li & Pollatsek, 2020). By doing so, CRM successfully simulates some important Chinese sentence processing features, such as the segmentation of Chinese overlapping ambiguous strings (Huang et al., 2021). However, it should be noted that CRM only simulated how strings of Chinese characters are activated and segmented, but left the question open that how identified words are integrated. The finding in the current study that words are integrated incrementally in Chinese reading implies that Chinese characters or words within a perceptual span are accessed in a parallel fashion, but are integrated in a serial and incremental fashion.

In Experiment 2, some native-like eye-movement patterns were observed in high-proficiency Dutch-Chinese learners. L2ers exhibited a significant plausibility

effect on Noun1 and a reversed plausibility effect on Noun2, which indicated that L2ers could automatically use the plausibility information of the Cl-Noun associations to build interpretations and solve syntactic ambiguities incrementally. However, it should be noted that L2ers' processing is not completely native-like. The significantly higher probability of RI in the Num+Cl region in the plausible condition than in the implausible condition, and the high RI in Noun1 region in both conditions indicated that when reanalyzing the nominal phrase structure, L2 participants not just regressed back to Noun1, they also regressed back to the Num+Cl region to resort to the Cl information. Furthermore, different from native speakers who showed a numerical tendency to a reverse plausibility effect on TT in Noun1 region, L2ers showed a significant plausibility effect on Noun1. This nonnative-like pattern implies that compared to native speakers, it took longer for L2ers to start the reanalysis procedure. In general, L2 participants exhibited native-like patterns in the early processing stage (reflected in GD and SKR), while they exhibited some nonnative-like patterns in the late processing stage (especially the reanalysis procedure, reflected in RI and TT). We speculate that L2 participants' native-like patterns in the early processing stage support the Feature Reassembly Hypothesis that late L2ers can eventually acquire unique-to-L2 knowledge and use it automatically in on-line processing, even though they need more time and information to resolve the temporary structure ambiguity.

As one of the main concerns of the current research was to explore the ultimate L2 acquisition of unique-to-L2 knowledge, we recruited Dutch-Chinese learners with a high level of Chinese proficiency. It would be interesting to further explore how

L2-Chinese learners' use of unique-to-L2 knowledge would develop along with their different Chinese proficiency levels. According to the Feature Reassembly Hypothesis, late L2ers' use of unique-to-L2 knowledge would improve with increased L2 proficiency. In other words, unlike high-proficiency Dutch-Chinese learners, beginning and intermediate L2-Chinese learners may not be able to make use of Cl-Noun association immediately during on-line processing. Furthermore, due to their short experience of learning Chinese, the uncertainty of each incoming Chinese character/word may lead them to use a delayed strategy instead of an immediate one in semantic integration. Thus, we speculate that beginning and intermediate L2-Chinese learners may not exhibit an immediate plausibility effect on Noun1, and consequently, no plausibility effect on Noun2. Future study is needed to test this speculation.

Conclusion

In conclusion, we found a plausibility effect in the initial constituent region and a reversed plausibility effect in the second constituent region in the on-line process of 4-character Chinese novel noun-noun combinations for both native Chinese speakers and high-proficiency Dutch-Chinese learners. These findings indicate that both native Chinese speakers and L2ers can immediately use the plausibility information of nouns in on-line processing. These results support the immediate account of the plausibility effect in Chinese processing and the Feature Reassembly Hypothesis in L2 acquisition.

Declarations

Conflicts of interest/competing interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Reference

- Aaronson, D., & Ferres, S. (1986). Sentence processing in Chinese-American bilinguals. *Journal of Memory and Language*, 25, 136–162.
- Adams, B. C., Clifton, C., & Mitchell, D. C. (1998). Lexical guidance in sentence processing? *Psychonomic Bulletin & Review*, 5, 265-270.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bates, D., Mächler, M., & Bolker, B. M. (2011). LME4: linear mixed-effects models using S4 classes. R Package Version 0.999375-39.
- Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, 48, 615-634.
- Chen, H. C. (1996). Chinese reading and comprehension: A cognitive psychology perspective. In M. H. Bond (Ed.), *Handbook of Chinese psychology* (pp. 43– 62). Hong Kong, China: Oxford University Press.
- Chen, H. C. (1999). How do readers of Chinese process words during reading for comprehension? In J. Wang, A. W. Inhoff, & H. C. Chen (Eds.), *Reading Chinese*

script: A cognitive analysis (pp. 257 – 278). Mahwah, NJ: Erlbaum.

Chen, H. C., Song, H., Lau, W. Y., Wong, K. F. E., & Tang, S. L. (2003).

Developmental characteristics of eye movements in reading Chinese. In C.

McBride-Chang & H.-C. Chen (Eds.), *Reading development in Chinese children*

(pp. 157–169). Westport, CT: Praeger.

Chen, L., Shu, H., Liu, Y. Y., & Zhao, J. J. (2007). ERP signatures of subject-verb

agreement in L2 learning. *Bilingualism: Language and Cognition*, 10,

161-174.

Crone, E. A., Richard Ridderinkhof, K., Worm, M., Somsen, R. J., & Van Der Molen,

M. W. (2004). Switching between spatial stimulus-response mappings: A

developmental study of cognitive flexibility. *Developmental science*, 7, 443-455.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path

sentences. *Journal of Memory and Language*, 30, 725-745.

Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis

and reanalysis. *Canadian Journal of Experimental Psychology*, 47, 247-275.

Frazier, L. & Clifton, C. (1998). Sentence reanalysis and visibility. In J.D. Fodor & F.

Ferreira (Eds). *Reanalysis in Sentence Processing*. Dordrecht, NL:Kluwer.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence

comprehension: Eye movements in the analysis of structurally ambiguous

sentences. *Cognitive Psychology*, 14, 178-210.

- Frenck-Mestre, C., Foucart, A., Carrasco-Ortiz, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language: Evidence from ERPs. *Eurosla yearbook*, 9, 76-106.
- Gajewski, P. D., Wild-Wall, N., Schapkin, S. A., Erdmann, U., Freude, G., & Falkenstein, M. (2010). Effects of aging and job demands on cognitive flexibility assessed by task switching. *Biological Psychology*, 85, 187-199.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369-397.
- Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, 36, 129-173.
- Huang, L., Staub, A., & Li, X. (2021). Prior context influences lexical competition when segmenting Chinese overlapping ambiguous strings. *Journal of Memory and Language*, 118, Article 104218.
- Hyönä, J., Pollatsek, A., Koski, M., & Olkonieni, H. (2020). An eye-tracking study of reading long and short novel and lexicalized compound words. *Journal of Eye Movement Research*, 13, 3-20.
- Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 20-34.
- Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles

- with definite plurals in L2 English. *Language learning*, 60, 877-925.
- Ionin, T., Baek, S., Kim, E., Ko, H., & Wexler, K. (2012). That's not so different from the: definite and demonstrative descriptions in second language acquisition. *Second language research*, 28, 69-101.
- Jackson, C. N. (2007). The use and non-use of semantic information, word order, and case marking during comprehension by L2 learners of German. *Modern Language Journal*, 91, 418-432.
- Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, 12, 65-82.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603-634.
- Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language acquisition. *Language Learning*, 57, 1-33.
- Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language learning*, 61, 940-967.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103, 54-69.

- Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language processing research. *Language Learning, 46*, 283-326.
- Kennison, S. M. (2005). Different time courses of integrative semantic processing for plural and singular nouns: Implications for theories of sentence processing. *Cognition, 95*, 269-294.
- Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language, 109*, 68-74.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*.
doi:10.18637/jss.v082.i13
- Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research, 25*, 173-227.
- Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice, 2*, 59-67.
- Lexicon of Common Words in Contemporary Chinese Research Team. (2008). *Lexicon of common words in contemporary Chinese*. Commercial Press.
- Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*. Advanced online. <http://dx.doi.org/10.1037/rev0000248>
- Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words

- during reading. *Cognitive Psychology*, 58, 525-552.
- Liu, P., Li, W., Lin, N., Li, X. (2013). Do Chinese readers follow the national standard rules for word segmentation during reading? *PLoS ONE*, 8(2), e55440.
- Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, 60, 154-193.
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eyetracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 940-961.
- Rayner, K. (2009). The Thirty-Fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457–1506.
- Rayner, K., Cook, A. E., Juhasz, B. J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements. *British Journal of Psychology*, 97, 467-482.
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16, 1-21.
- Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, 32, 299-331.

Slabakova, R. (2016). *Second language acquisition*. Oxford, UK: Oxford University Press.

Slatterly, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013).

Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69, 104-120.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 33, 550-569.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time

course of plausibility effects on eye movements in reading: Evidence from noun–noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1162-1169.

Sturt, P., Pickering, M. J., & Crocker, M.W. (1999). Structural change and reanalysis

difficulty in language comprehension. *Journal of Memory and Language*, 40, 136-150.

Wang, S., Chen, H. C., Yang, J., & Mo, L. (2008). Immediacy of integration in

discourse comprehension: Evidence from Chinese readers' eye movements. *Language and Cognitive Processes*, 23, 241-257.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge,

UK: Cambridge University Press.

Williams, J. (2006). Incremental interpretation in second language sentence

processing. *Bilingualism: Language and Cognition*, 9, 71-88.

- Williams, J., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22, 509-540.
- Yan, M., Zhou, W., Shu, H., & Kliegl R. (2015). Perceptual span depends on font size during the reading of Chinese sentence. *Journal of Experimental Psychology: Learning Memory and Cognition*, 41, 209-219.
- Yang, J. M., Wang, S. P., Chen, H. C., & Rayner, K. (2009). The time course of semantic and syntactic processing in Chinese sentence comprehension: Evidence from eye movements. *Memory & Cognition*, 37, 1164-1176.
- Yang, J., Staub, A., Li, N., Wang, S., & Rayner, K. (2012). Plausibility effects when reading one- and 2-character words in Chinese: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1801-1809.
- Yao, P., & Chen, B. (2016). Cross-linguistic differences affect late Chinese-English learners on-line processing of English tense and aspect. *International Journal of Bilingualism*, 21, 268-290.
- Zhou, J., & Li, X. (2021). On the segmentation of Chinese incremental words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <https://doi.org/10.1037/xlm0000984>

Table 1.*The Descriptive Features of Noun1 in Two Conditions*

	Plausible	Implausible	<i>p</i>
Frequency (occurrence per million)	15.89	19.59	.45
No. of strokes of 1 st character	8.13	7.93	.81
No. of strokes of 2 nd character	9.56	8.33	.18
Mean No. of strokes	17.7	16.26	.27
Bigram frequency of Noun1+Noun2 (occurrence per thousand)	0.057	0.034	.45
Norming score of the nominal phrases	4.61	4.67	.54
Norming score of the Classifier-Noun1 association	4.85	1.12	<.001
Norming score of the Noun1+Noun2 combinations	4.65	4.74	.28

Table 2

Mean and SE (in parentheses) for Each Index in Each Region for Native Chinese Speakers & Dutch-Chinese Learners

	Numeral+Classifier		Noun1		Noun2	
	Plausible	Implausible	Plausible	Implausible	Plausible	Implausible
<i>Native Chinese speakers (L1ers)</i>						
FFD	208 (5.43)	204 (5.88)	268 (6.01)	267 (6.73)	261 (5.15)	263 (5.91)
GD	225 (8.35)	216 (8.28)	302 (7.70)	329 (11.30)	304 (7.13)	306 (7.88)
GP	236 (12.94)	237 (10.93)	343 (15.69)	379 (17.77)	488 (21.58)	417 (22.41)
TT	319 (18.13)	299 (15.71)	551 (28.89)	533 (26.06)	581 (23.75)	486 (23.99)
SKR	.49 (.03)	.50 (.03)	.28 (.03)	.23 (.03)	.18 (.02)	.22 (.02)
RI	.31 (.03)	.28 (.02)	.49 (.03)	.39 (.03)	.22 (.02)	.21 (.02)
RO	.05 (.01)	.03 (.01)	.09 (.01)	.11 (.02)	.25 (.02)	.18 (.02)
<i>Dutch-Chinese learners (L2ers)</i>						
FFD	212 (7.08)	217 (8.70)	271 (7.53)	279 (8.27)	267 (6.62)	268 (8.09)
GD	232 (8.48)	232 (10.47)	322 (11.06)	353 (14.38)	318 (9.10)	319 (11.03)
GP	257 (21.76)	260 (18.32)	355 (23.66)	402 (22.88)	509 (31.06)	431 (34.33)
TT	348 (24.24)	342 (20.15)	577 (33.07)	648 (39.60)	623 (38.87)	526 (40.11)
SKR	.26 (.02)	.23 (.02)	.19 (.03)	.13 (.02)	.17 (.02)	.19 (.02)
RI	.43 (.02)	.33 (.02)	.43 (.03)	.45 (.04)	.30 (.03)	.28 (.03)
RO	.05 (.01)	.05 (.01)	.14 (.02)	.18 (.02)	.32 (.02)	.23 (.02)

Note. The data in bold are those with significant differences between plausible and implausible conditions. FFD stands for first fixation duration, GD stands for gaze duration, GP stands for go-past time, TT stands for total reading time, SKR stands for skipping rate, RI stands for regression-in probability, and RO stands for regression-out probability.

Table 3*The Results of LME/GLME Models in Three Regions for Native Speakers*

	Estimate	SE	t/z	p
Numeral+Classifier Region				
<i>First fixation duration</i>				
(Intercept)	209.91	7.76	27.06	<.001
Plausibility	-4.16	8.53	-0.49	.63
<i>Gaze duration</i>				
(Intercept)	230.16	11.25	20.45	<.001
Plausibility	-12.02	12.05	-1.00	.32
<i>Go past time</i>				
(Intercept)	242.21	13.17	18.39	<.001
Plausibility	-1.18	14.71	-0.08	.94
<i>Total reading time</i>				
(Intercept)	322.77	19.93	16.20	<.001
Plausibility	-22.93	18.15	-1.26	.21
<i>Skipping rate</i>				
(Intercept)	-0.05	0.14	-0.34	.74
Plausibility	0.04	0.13	0.30	.77
<i>Regression in probability</i>				
(Intercept)	-0.94	0.17	-5.64	<.001
Plausibility	-0.21	0.17	-1.21	.23
<i>Regression out probability</i>				
(Intercept)	-3.40	0.27	-12.78	<.001
Plausibility	-0.34	0.30	-1.13	.26
Noun1 Region				
<i>First fixation duration</i>				
(Intercept)	262.97	7.41	35.49	<.001
Plausibility	6.40	7.94	0.81	.42
<i>Gaze duration</i>				
(Intercept)	299.63	11.42	26.25	<.001
Plausibility	31.88	12.51	2.55	.01
<i>Go past time</i>				
(Intercept)	343.52	20.61	16.66	<.001
Plausibility	40.16	22.16	1.81	.08
<i>Total reading time</i>				
(Intercept)	555.83	34.53	16.10	<.001
Plausibility	-22.71	34.16	-0.67	.51
<i>Skipping rate</i>				
(Intercept)	-1.36	0.23	-5.94	<.001
Plausibility	-0.41	0.17	-2.42	.02

<i>Regression in probability</i>				
(Intercept)	-0.09	0.18	-0.49	.62
Plausibility	-0.44	0.18	-2.45	.01
<i>Regression out probability</i>				
(Intercept)	-2.73	0.24	-11.27	<.001
Plausibility	0.14	0.26	0.55	.58
Noun2 Region				
<i>First fixation duration</i>				
(Intercept)	261.79	6.17	42.42	<.001
Plausibility	-0.52	6.93	-0.08	.94
<i>Gaze duration</i>				
(Intercept)	304.31	10.47	29.06	<.001
Plausibility	-0.54	13.17	-0.04	.97
<i>Go past time</i>				
(Intercept)	488.14	28.72	16.99	<.001
Plausibility	-73.37	34.00	-2.16	.04
<i>Total reading time</i>				
(Intercept)	579.96	30.02	19.32	<.001
Plausibility	-94.08	30.77	-3.06	<.001
<i>Skipping rate</i>				
(Intercept)	-1.77	0.17	-10.28	<.001
Plausibility	0.24	0.18	1.35	.18
<i>Regression in probability</i>				
(Intercept)	-1.46	0.17	-8.64	<.001
Plausibility	-0.04	0.20	-0.22	.82
<i>Regression out probability</i>				
(Intercept)	-1.31	0.18	-7.33	<.001
Plausibility	-0.46	0.22	-2.15	.03

Table 4*The Results of LME/GLME Models in Three Regions for Dutch-Chinese learners*

	Estimate	SE	t/z	p
Numeral+Classifier Region				
<i>First fixation duration</i>				
(Intercept)	212.35	7.96	26.68	<.001
Plausibility	6.86	6.84	1.00	.32
<i>Gaze duration</i>				
(Intercept)	231.44	9.96	23.23	<.001
Plausibility	-0.25	8.00	-0.03	.98
<i>Go past time</i>				
(Intercept)	262.25	19.01	13.80	<.001
Plausibility	-4.85	18.45	-0.26	.79
<i>Total reading time</i>				
(Intercept)	350.47	23.39	14.98	<.001
Plausibility	-8.32	18.48	-0.45	.65
<i>Skipping rate</i>				
(Intercept)	-1.07	0.12	-8.86	<.001
Plausibility	-0.18	0.17	-1.04	.30
<i>Regression in probability</i>				
(Intercept)	-0.30	0.12	-2.52	.01
Plausibility	-0.43	0.16	-2.67	<.01
<i>Regression out probability</i>				
(Intercept)	-2.96	0.24	-12.21	<.001
Plausibility	-0.10	0.31	-0.33	.74
Noun1 Region				
<i>First fixation duration</i>				
(Intercept)	271.64	7.71	35.25	<.001
Plausibility	9.47	7.54	1.26	.21
<i>Gaze duration</i>				
(Intercept)	323.94	12.68	25.55	<.001
Plausibility	31.84	12.90	2.47	.02
<i>Go past time</i>				
(Intercept)	352.73	26.10	13.51	<.001
Plausibility	53.68	26.80	2.00	.05
<i>Total reading time</i>				
(Intercept)	576.65	36.80	15.67	<.001
Plausibility	76.61	27.32	2.80	<.01
<i>Skipping rate</i>				
(Intercept)	-1.68	0.22	-7.68	<.001
Plausibility	-0.48	0.22	-2.12	.03

<i>Regression in probability</i>				
(Intercept)	-0.36	0.18	-1.99	.05
Plausibility	0.14	0.20	0.70	.48
<i>Regression out probability</i>				
(Intercept)	-1.94	0.19	-10.15	<.001
Plausibility	0.31	0.23	1.34	.18
Noun2 Region				
<i>First fixation duration</i>				
(Intercept)	266.94	7.35	36.33	<.001
Plausibility	0.07	7.85	0.01	.99
<i>Gaze duration</i>				
(Intercept)	318.11	11.57	27.49	<.001
Plausibility	-0.38	13.87	-0.03	.98
<i>Go past time</i>				
(Intercept)	513.51	34.76	14.77	<.001
Plausibility	-81.29	37.09	-2.19	.03
<i>Total reading time</i>				
(Intercept)	622.32	43.71	14.24	<.001
Plausibility	-96.95	35.68	-2.72	<.01
<i>Skipping rate</i>				
(Intercept)	-1.73	0.19	-9.13	<.001
Plausibility	0.10	0.21	0.50	.62
<i>Regression in probability</i>				
(Intercept)	-0.96	0.17	-5.51	<.001
Plausibility	-0.09	0.23	-0.39	.70
<i>Regression out probability</i>				
(Intercept)	-0.77	0.10	-7.43	<.001
Plausibility	-0.47	0.16	-3.02	<.01