"Nage" Before vs. After Relative Clauses in Chinese: A Processing View*

James H-Y. Tai  Pei-fen Du
National Chung Cheng University

Abstract Chao (1968) proposes that the order of *Nage* before vs. after determines whether there are restrictive or descriptive uses of DE in Chinese relative clauses (RCs). *Nage* in post-RC is for restrictive use, whereas *Nage* in pre-RCs is for descriptive use. Recently Sheng and Wu (2012) have shown that *Nage* prefers to occur before subject-extracted relative clauses (SRCs) and after object-extracted relative clause (ORCs) in their spoken corpus and sentence production experiment. They attribute the distribution to pragmatic functions. This paper proposes that an effect of canonicity of thematic order-AGENT-ACTION-PATIENT may also affect real-time processing (Small et al. 2000, Lin 2012). Based on this effect, the structures of *Nage* in pre-SRCs and post-ORCs conform to the canonical thematic order, whereas the structures of *Nage* in post-SRCs and pre-ORCs do not. A self-paced reading experiment was conducted, and the result shows that *Nage* in pre-SRCs is easier to process than *Nage* in post-SRCs, whereas *Nage* in post-ORCs is easier to process than *Nage* in pre-ORCs. It suggests that when sentences conform to the canonical thematic order, it facilitates to process the regions of RC and head noun, and vice versa. The main processing difference for *Nage* position was interpreted as a processing demand for readers, and does not rely on the pragmatic functions only.

* This study was in part supported by grants from the National Science Council of Taiwan (NSC-101-2410-H194-115-MY3). We are in debt to Prof. James Myers, who advised us of experimental designs and statistical analysis. We have also benefited from suggestions and comments from research assistants in Language Processing Lab, National Chung Cheng University: Chia-wen Lo, Guo-ming Hong, Yu-rui Liu, Yu-guang Ko, and Chen-tsung Yang. We thank all subjects who participated in the experiments. We are, of course, solely responsible for all possible errors herein.
1. Introduction

In Chinese, the combination of demonstrative + (number) + classifier (DCL) can either precede or follow Chinese relative clauses (RCs), as in (1) and (2). In the context of discussing restrictive and descriptive uses of de, Chao (1968: 286) analyzed (1) as with descriptive use and (2) as with restrictive use.

(1) 那位 [戴 眼鏡 的] 先生 是 誰?
nawei dai yanjing de xiansheng shi shei
DCL wore glasses DE gentleman is who
“Who is that gentleman (who incidentally is) wearing glasses?”

(2) [戴 眼鏡 的] 那位 先生 是 誰?
dai yanjing de nawei xiansheng shi shei
wore glasses DE DCL gentleman is who
“Who is the gentleman who is wearing glasses?”

In (1), the DCL nawei precedes the relative clause, which is a descriptive use in that nawei dai yanjing de only describe the xiansheng rather than distinguish it from other xiansheng. By contrast, in (2), the DCL nawei follows the relative clause, which has a restrictive use in that xiansheng is referring to xiansheng who is wearing glasses only rather than other xiansheng. At the same time, Chao (1968: 286) observed that if a stress is placed on the modifier dai yanjing in (1), the sentence will also have the restrictive sense as in (2). Later Hashimoto (1971) and Tang (1979) have analyzed this difference as equivalent to the distinction between restrictive and non-restrictive relative clauses (RCs) in English. Nevertheless, Teng (1987) and Lin (2003) and others show that Chinese doesn’t have the distinction, all RCs being restrictive.

Recently, Huang, Li and Li (2009) observe that there are three positions that a Chinese RC may appear in a nominal phrase, as shown in (3).

(3) Huang, Li and Li (2009: 214)

Demonstrative + (Number) + Classifier + Noun

\[ \begin{array}{c}
\text{Demonstrative} \\
\text{(Number)} \\
\text{Classifier} \\
\text{Noun}
\end{array} \]

\[ \begin{array}{c}
\text{position I} \\
\text{position II} \\
\text{position III}
\end{array} \]

Chinese RCs in position I appear before the demonstrative, as in (4).

---

1 In Chao’s work, classifiers are individual measures. A measure, as a bound morpheme, forms a D-M compound with the demonstrative (p.584). Thus D-M is used in Chao (1968) rather than DCL (demonstrative + classifier). Also the English translations of (1) and (2) sentences are Chao’s original translations to show the meaning contrast.
(4) Huang, Li and Li (2009: 214)

[[他 喜歡 的 RC] 那 (一)個 孩子 NP]
ta xihuan de na (yi)-ge haizi
he like DE that one-CL child
“the child that he likes”

Chinese RCs in position II appear between a demonstrative and a number, as in (5).

(5) Huang, Li and Li (2009: 215)

[這 [張三 說 出來 的 RC] 一句 話 NP] 比
zhe Zhangsan shuo chulai de yiju hua bi
this Zhangsan speak out DE one-CL word compare

[那 [李四 寫 出來 的 RC] 一百句 話 NP] 還 有用
na Lisi xie chulai de yibaiju hua hai youyong
that Lisi write out DE one.hundred-CL word more useful
“This one sentence that Zhangsan uttered is even more useful than those hundred sentences that Lisi wrote.”

Huang et al. (2009) suggest that the position II is the most unnatural position and most complex example for Chinese RCs. Chinese RCs in position III occur between a classifier and a noun as (6).

(6) Huang, Li and Li (2009: 215)

[那 (一) 個 [他 喜歡 的 RC] 孩子 NP]
na yi ge ta xihuan de haizi
that one-CL he like DE child
“the child that he likes”

For the purpose of this study, we focus on the distinction between position 1 and position III. Huang (1982) suggests that the two positions involve different scopes of modification, and function differently. In position I, the demonstrative is a deictic expression, referring to a designated entity (“that one”). In position III, the demonstrative is an anaphoric expression, identified by the preceding relative. A similar distinction between a descriptive versus identificational use of the two positions is also proposed by Lu (1998).

Incidentally, Zhang (2006) holds that there are two types of Chinese indefinite nominals: (i) the inner modifier nominal (IMN) and (ii) the outer modifier nominal (OMN) in accordance with the positions of a modifier of nominal such as a PP, AP, NP and RC modifier. IMN refers to the order of numeral-classifier-modifier-de-N that the modifier occurs internal position of the nominal (e.g., san ge [dai yanjing de] xuesheng “three students who wear glasses”). In contrast, OMN refers to the order of modifier-de-numeral-classifier-N that the modifier occurs the left-peripheral position.
of the nominal (e.g., [dai yanjing de} san ge xuesheng “three students who wear glasses”). It should be noted here that the present study focuses on the position of Nage that can either precede or follow a RC (i.e. position III and I), and does not address Zhang’s (2006) analysis of the positions of a modifier of nominal in general.

Previous corpus-based research (Wu 2009, Ming 2010) has shown that Nage in pre-SRCs (subject extracted relative clauses) outnumber Nage in post-SRCs, whereas Nage in post-ORCs (object extracted relative clauses) outnumber Nage in pre-ORCs. This asymmetric distribution was also reported by a combination of corpus-based and experimental study (Sheng and Wu 2012). They employed a spoken corpus and sentence-production experiment to examine the distribution of DCL Nage in SRCs and ORCs. They found an asymmetric distribution pattern according to 357 RCs investigated from a Chinese live TV-show: (i) Nage tended to occur before SRCs, and (ii) Nage tended to occur after ORCs. Furthermore, in their sentence-production experiment, sentences were chunked into 4 parts (DCL/ RC/ head noun/ matrix clause), and each randomly assigned to 4 boxes in a diamond layout on a visual display. The participants produced the utterance based on these four parts of the sentences. The results also show that the participants have preference to produce Nage in pre-SRC (Nage-SRCs), as compared to Nage in post-SRCs (SRCs-Nage), whereas they have preference to produce Nage in post-ORCs (ORCs-Nage), as compared to Nage in pre-ORCs (Nage-ORCs) (p < .0001). Their finding was interpreted within the Audience Design Model to the effect that speakers favor to use Nage before SRCs as an early cue for their listeners (or audience) to ease comprehension; but to use ORCs-Nage to avoid structural ambiguity with garden path in Nage-ORCs. The example (7) with structural ambiguity is extracted from Sheng and Wu (2012).

(7) Nage-ORCs:

那個 星探 遇見 的 女孩 長得 很 漂亮
nage xingtan yujian de nuhai zhangde hen piaoliang
DCL star-hunter meet DE girl look very beautiful
“The girl whom the star-hunter met looks very beautiful.”

They suggested that Nage can be compatible with both the embedded noun xingtan “star-hunter” and head noun nuhai “girl” when Nage occurs in initial position of ORCs, and this structure is unable to clearly express whether speakers refer to either the embedded noun or head noun for their listeners. Hence, it resulted in reducing to use this type of RC sentences for speakers.

Lin (2012) observed an effect of thematic order in discourse context when processing Chinese SRCs and ORCs. Since Chinese is viewed as SVO word order with the canonical thematic order like AGENT-ACTION-PATIENT. The discourse with canonical condition is given in (8), and the target sentence either ORCs or SRCs is also given in (9).
(8) **Discourse Context with canonical condition** (Lin 2012: 45)

一棟 公寓 裡 住了 房東 以及 兩個 房客
yidong gongyu li zhule fangdong yiji liangge fangke
one apartment in lived landlord and two tenants
“A landlord and two tenants lived in an apartment.”

一位 住戶AGENT 噪音 太 大 吵醒了ACTION 房東PATIENT
yiwei zhuhu zaoyin tai da chaoxingle fangdong
one tenant noise too loud woke up landlord
“One of the tenants was too noisy and woke up the landlord.”

這位 房東AGENT 噪音 也 不 小 吵醒了ACTION 另一位 住戶 PATIENT
zhewei fangdong zaoyin ye bu xiao chaoxingle lingyiwei zhuhu
this landlord noise also not small woke up the other tenant
“This landlord also made noises and woken up the other tenant.”

小明: 我 聽說 其中 一名 住戶 被 微軟 雇用了
Xiaoming: wo tingshuo qizhong y1mmg zhuhu bei Weiruan guyongle
“I heard among them one tenant BEI Microsoft hired

哪一位 住戶 被 微軟 雇用?
nayiwei zhuhu bei Weiruan guyong
“Which tenant was hired by Microsoft?”

(9) **Target Sentences** (Lin 2012: 46)

a. **ORC:**

[[房東 AGENT 吵醒 ACTION 的] 住戶 PATIENT 被 微軟 雇用]
Fangdong chaoxing de zhuhu bei Microsoft guyong
landlord woke up DE tenant BEI Microsoft hire
“The tenant who the landlord woke up was hired by Microsoft.”

b. **SRC:**

[[吵醒 ACTION 房東 PATIENT 的] 住戶 AGENT 被 微軟 雇用]
chaoxing fangdong de zhuhu bei Microsoft guyong
woke up landlord DE tenant BEI Microsoft hire
“The tenant who woke up the landlord was hired by IBM.”

Lin (2012) emphasized that not only structural frequency information associated with lexical item but also the linear order of thematic roles play important roles. He conducted two self-paced reading experiments and successfully predicted that the target sentence (9a) for ORCs were read faster in the first and second regions after the head noun when the discourse context (8) given fully overlapped thematic orders-AGENT-ACTION-PATIENT (see the bold words).

Accordingly, we suggest that *Nage* constructions also imply the effect of the thematic order, as in (10) and (11). For SRCs, (10a) conforms to the canonical
thematic order (AGENT-ACTION-PATIENT), but (10b) does not. For ORCs, (11b) conforms to the canonical thematic order, but (11a) does not.

(10)a. **Nage-SRC with canonical thematic order:**

那個**AGENT**[上午 ** ACTION** 讀者**PATIENT** 的] 作家
nage shangwu chengzan duzhe de zuojia
DCL morning praised readers DE writer
“the writer who praised the readers this morning”

b. **SRC-Nage with non-canonical thematic order:**

[上午 ** ACTION** 讀者**PATIENT** 的] 那個 作家**AGENT**
shangwu chengzan duzhe de nage zuojia
morning praised readers DE DCL writer
“the writer who the readers praised this morning”

(11)a. **Nage-ORC with non-canonical thematic order:**

那個**AGENT**[上午 讀者**AGENT** ** ACTION** 的] 作家**PATIENT**
nage shangwu duzhe chengzan de zuojia
DCL morning readers praised DE writer
“the writer who the readers praised this morning”

b. **ORC-Nage with canonical thematic order:**

[上午 讀者**AGENT** ** ACTION** 的] 那個 作家**PATIENT**
shangwu duzhe chengzan de nage zuojia
morning readers praised DE DCL writer
“the writer who the readers praised this morning”

Hence, in addition to the discourse account from the production experiment (Sheng and Wu, 2012), we also propose that there is an effect of canonicity of thematic order in real-time processing.

While Sheng and Wu (2012) have analyzed the distribution and functions of Chinese *Nage* from the corpus-based studies as well as production experiment, real-time data from comprehension experiment are still needed to understand the role of *Nage* in processing of RC in Chinese. The goal of this study is to see if the asymmetric pattern observed in the production experiments and the effect of canonicity of thematic order also holds in comprehension experiments.

2. **Self-paced reading experiment**

For the comprehension experiment, we employed a self-paced reading task to examine how the distribution of Chinese DCL *Nage* in SRCs and ORCs affects real-time processing. Beyond the frequency effect from the corpus analysis (Sheng and Wu 2012), we also want to know how the canonicity of thematic order (i.e. default semantic/syntactic linking in Chinese), affect on-line parsing. We predict that
the Chinese RCs which are compatible with the canonical thematic order (i.e. Nage-SRCs and ORCs-Nage) may facilitate sentence processing, particularly at the regions of RC and head noun, and vice versa.

2.1 Method

2.1.1 Participants

Twenty undergraduate and graduate students (11 males, 9 females) from National Chung Cheng University participated in this experiment. All participants were over the age of 18, and the age range was from 18 to 27 ($M = 24.5, SD = 2.72$). They had normal or corrected-to-normal vision by self-report. All participants were Taiwan Mandarin native speakers, while all were exposed to Taiwan Southern Min or Hakka since birth. They also reported no speech, hearing, language disorders, and brain injury. They each received a gift for participating in the experiment.

2.1.2 Materials

Twenty sets of target sentences were manipulated by two factors: (i) Extraction Types (SRCs vs. ORCs) and (ii) Position of Nage, before RC (Nage-RC) vs. after RC (RC-Nage), which denotes either Nage precedes or follows the RCs. A total of 80 ($= 20$ sets $\times 2$ Extraction Types $\times 2$ Positions of Nage) were designed, and one of the sets for target sentences is given in Table 1.

<table>
<thead>
<tr>
<th>Sentence types</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nage-SRC</td>
<td>常早幫助村民的人農夫問候了村長</td>
</tr>
<tr>
<td></td>
<td>DCL ADV RV RO DE MS MV MO</td>
</tr>
<tr>
<td></td>
<td>Nage shangwu bangzhu cunmin de nongfu wenhoule cunzhang DCL morning help villager DE farmer greeted village head</td>
</tr>
<tr>
<td></td>
<td>&quot;The farmer who helped the villager greeted the village head this morning.&quot;</td>
</tr>
<tr>
<td>SRC-Nage</td>
<td>[上午幫助村民的] 那個農夫問候了村長</td>
</tr>
<tr>
<td></td>
<td>ADV RV RO DE DCL MS MV MO</td>
</tr>
<tr>
<td></td>
<td>Shangwu bangzhu cunmin de nage nongfu wenhoule cunzhang morning help villager DE DCL farmer greeted village head</td>
</tr>
<tr>
<td></td>
<td>&quot;The farmer who helped the villager greeted the village head this morning.&quot;</td>
</tr>
<tr>
<td>Nage-ORC</td>
<td>[上午村民幫助的]農夫問候了村長</td>
</tr>
<tr>
<td></td>
<td>DCL ADV RS RV DE MS MV MO</td>
</tr>
<tr>
<td></td>
<td>Nage shangwu cunmin bangzhu de nongfu wenhoule cunzhang DCL morning villager help DE farmer greeted village head</td>
</tr>
<tr>
<td></td>
<td>&quot;The farmer who the villager helped greeted the village head this morning.&quot;</td>
</tr>
<tr>
<td>ORC-Nage</td>
<td>[上午村民幫助的]那個農夫問候了村長</td>
</tr>
<tr>
<td></td>
<td>ADV RS RV DE DCL MS MV MO</td>
</tr>
</tbody>
</table>
|                | Shangwu cunmin bangzhu de nage nongfu wenhoule cunzhang
The farmer who the villager helped greeted the village head this morning.

Note that in Table 1, an extra temporal adverbial such as shangwu “morning” was attached to the initial position of RCs to avoid the garden-path effect. Without a temporal adverbial in Nage-ORC, the sentence will be interpreted as Nage cunmin “that villager” in the initial stage of processing rather than the intended correct interpretation Nage nongfu “that farmer”.

In order to prevent participants from noticing the patterns of experimental stimuli, forty filler sentences, which were not related to the intent of the experiment, were included. These filler sentences consist of diverse sentence types selected from Academia Sinica Balanced Corpus of Modern Chinese version 4.0 (Sinica Corpus) compiled by Academia Sinica in Taiwan (CKIP1995/1998). In addition, half of the filler items contain a possessive marker DE to distract participants’ attention. The length of filler sentences was similar to the length of target sentences consisted of 7 to 9 words.

Overall, twenty sets of target sentences and forty filler sentences were counterbalanced across four conditions for twenty participants, and thus, each participant read a total of sixty (= 20+40) sentences for each version.

2.1.3 Procedure

A self-paced reading experiment was conducted with Linger version 2.94 software developed at MIT (Rohde 2003). This technique provides reliable online measures of processing difficulty or speed. Participants performed the experiment individually in a sound-proof cubicle room. At the beginning of the experiment, the written instruction was first given on the computer screen, and the experimenter gave the oral instruction to them later. Four practice trials were provided to familiarize the participants with the task before the formal test session. For each trial, a fixation symbol “+”, which is an impending start of a sentence at the center of the computer screen. A word or phrase appeared in the center of computer screen when the participants pressed the spacebar to control the presentation of the successive chunks of the sentence. In this paradigm, while each word or phrase appears, the preceding word disappears. Hence, the times for each chunk of sentences that the participants read were recorded.

To ensure the participants were paying attention in this task, they were asked to give an answer to a comprehension question after reading each sentence. In a yes-no question, the participants were instructed to hit the key “F” for true and the key “J” for false. Feedback was immediately provided after they answer the comprehension questions. All sentences were presented randomly in order to control for the possible effects of the order of presentation, such as fatigue or habituation. Each experimental stimulus was presented one at a time in randomized order. In addition, participants were asked to read the sentence at their own and natural pace, so that there were no time constraints for participants to read sentences. The whole experiment took approximately twenty minutes to finish.
2.2 Results

2.2.1 Comprehension-question accuracy

The grand mean accuracy rate of comprehension questions for all target trials was 88.5%. The mean accuracy rates across two main effects are shown in Figure 1. The comparison of Extraction Type (i.e. SRCs and ORCs) shows that Nage-SRCs were comprehended more accurately than SRCs-Nage, whereas ORCs-Nage was comprehended more accurately than Nage-ORCs. In short, a ranking hierarchy of the grand mean accuracy for each sentence type was:

Nage-SRCs > ORCs-Nage > SRCs-Nage = Nage-ORCs

![Accuracy for Comprehension Questions](image)

**Figure 1** The comparison of mean accuracy rates between SRCs and ORCs

To obtain maximized power of statistical analysis, a mixed-effects logistic regression (Baayen 2008, Pinheiro and Bates 2000) in the free statistics program R (R Development Core Team 2011) was used to analyze two main effects (Extraction Types, and Position of Nage) for the binomial response of accuracy. The likelihood ratio test first carried out the comparison between the by-subject-and-item model and the by-subject-only model.

```r
> acc.analysis.s = lmer (Accuracy ~ ExtractionType*PositionNage + (1|Subject),
family = "binomial", data = acc)
> acc.analysis.si = lmer (Accuracy ~ ExtractionType*PositionNage + (1|Subject) + (1|Item), family = "binomial", data=acc)
> anova (acc.analysis.s, acc.analysis.si)
```

Table 2 illustrates that accuracy has no significant difference either on Extraction Types ($p > .05$) or Position of Nage ($p > .05$). In addition, no significant interaction was observed ($p > .05$).
Table 2 Linear mixed-effects logistic regression model of the predictors for Accuracy

| Test Model                  | Predictors in mixed-effects logistic regression | Estimate | Std. Error | z value | Pr(>|z|) |
|-----------------------------|------------------------------------------------|----------|------------|---------|---------|
| Subject & Item              | (Intercept)                                     | 2.30259  | 0.24823    | 9.276   | 0.000***|
|                             | Extraction Type                                | 0.05352  | 0.22488    | 0.238   | 0.812   |
|                             | NagePosition                                    | 0.07501  | 0.22488    | 0.334   | 0.739   |
|                             | Extraction Type: NagePosition                   | 0.25087  | 0.22488    | 1.116   | 0.265   |

Note: “***”: p < .001

2.2.2 Reading time data

2.2.2.1 Interpretation of the measures

The main analysis for reading time was conducted by using a linear mixed-effects model in lme4 package (Bates, Maechler and Bolker 2011) in R (R Development Core Team 2012). Before analyzing the reading times data, the outliers for all reading time across different regions first were examined. The statistical outliers in this experiment were eliminated by using the method proposed by Hofmeister (2007) and Wu et al. (2012). Hofmeister (2007) proposes that the extremely high reading times reflect the less concentration on comprehending, parsing errors making, participants’ physical problems, or even the irrelevant reasons to cause the slowdown of online parsing. To remove these statistical outliers, we used the trim methods by proposed Wu et al. (2012). They recommended that reading times which are longer than 4,000 ms for a double-syllable Chinese word should be excluded. Moreover, Hofmeister (2007) suggest that the number of the removal of extreme outliers for self-paced reading task typically should be less than 3% of the data. Accordingly, 5 extreme outliers for all reading times (3200 data) we analyzed were removed, and the result fitted the standard that was 0.16% (< 3%). After elimination of the statistical outliers, the remaining reading times were log transformed (i.e. logarithm) in this experiment because the previous studies suggest several advantages to analysis log-transformed reading times instead of raw reading times (Jaeger, Gillespie and Graff 2010, Wu et al. 2012). First, it has the advantages of log transformation to correct and remove skewness because individual RT varies with different participants’ reading speeds as well as the difference of word length. Second, the data were analyzed using a LME predicted under a linear regression model, and hence, log transformation can reduce non-linear relations.

2.2.2.2 Analysis for each region

Four regions were analyzed across four sentence types, as shown Table 3. The gray parts in Table 3 stand for the regions in RCs: RV, RS, and RO. By contrast, the white parts stand for the regions in main clauses: DE and MS.

Table 3 Analysis of four regions across four sentence types

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nage_SRCs</td>
<td>RV</td>
</tr>
<tr>
<td></td>
<td>幫助 bangzhu</td>
</tr>
<tr>
<td>SRCs_Nage</td>
<td>RV</td>
</tr>
<tr>
<td></td>
<td>幫助 bangzhu</td>
</tr>
</tbody>
</table>
“Nage” Before vs. After Relative Clauses in Chinese: A Processing View / 285

<table>
<thead>
<tr>
<th>Nage ORCs</th>
<th>RS</th>
<th>RV</th>
<th>DE</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>村民 cunmin</td>
<td>幫助 bangzhù</td>
<td>的</td>
<td>冬夫 nongfu</td>
<td></td>
</tr>
<tr>
<td>ORCs_Nage</td>
<td>RS</td>
<td>RV</td>
<td>DE</td>
<td>MS</td>
</tr>
<tr>
<td>村民 cunmin</td>
<td>幫助 bangzhù</td>
<td>的</td>
<td>冬夫 nongfu</td>
<td></td>
</tr>
</tbody>
</table>

Note: “RV” = verbs in relative clauses; “RO” = objects in relative clauses; “RS” = subjects in relative clauses; “DE” = relativizers in matrix clause; “MS” = matrix subjects

Overall, a ranking hierarchy of the grand mean log reading times per region across four sentence types was:

Nage-ORCs > SRCs-Nage > Nage-SRCs > ORCs-Nage
2.723 > 2.706 > 2.702 > 2.690 (log reading times)

The result shows that ORCs-Nage were read faster than Nage-ORCs (2.690 vs. 2.723 log reading times); by contrast, Nage-SRCs were read faster than SRCs-Nage (2.702 vs. 2.706 log reading times).

Figure 2 illustrates the mean log reading times for four regions across four sentence types (i.e. Nage-SRCs, SRC-Nage, Nage-ORC and ORC-Nage).

![Figure 2 Mean log reading times for four regions across four sentence types](image)

Note: ***p < .001; *p < .05; “Nage Position*” stands for a significant effect of Position of Nage; “ET: PN*” stands for a significant interaction between Extraction Types (ET) and Position of Nage (PN).

Furthermore, two main effects (i.e. Extraction Types and Position of Nage) and their interaction were examined by using a linear mixed-effects model with the lme4
The likelihood ratio test first carried out the comparison between the by-subject-and-item model and the by-subject-only model for each region. The results showed that the by-subject-only model for each region fitted the data better than the by-subject-and-item model, as shown in Table 4. The overall statistical analysis for each region consists of the test model, estimate, standard error, t value, and pMCMC is reported in Table 4.

Table 4 Linear mixed-effects regression model of the predictors for four regions

<table>
<thead>
<tr>
<th>Region</th>
<th>Test Model</th>
<th>Predictors in LME</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>pMCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV/RS</td>
<td>Subject</td>
<td>(Intercept)</td>
<td>2.705062</td>
<td>0.031365</td>
<td>86.24</td>
<td>0.0001***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type</td>
<td>-0.001433</td>
<td>0.007648</td>
<td>-0.19</td>
<td>0.8658</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nage Position</td>
<td>0.014271</td>
<td>0.007648</td>
<td>1.87</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type: NagePosition</td>
<td>-0.02138</td>
<td>0.007648</td>
<td>-2.8</td>
<td>0.0046**</td>
</tr>
<tr>
<td>RO/RV</td>
<td>Subject</td>
<td>(Intercept)</td>
<td>2.702718</td>
<td>0.033332</td>
<td>81.08</td>
<td>0.0001***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type</td>
<td>-0.001193</td>
<td>0.006677</td>
<td>-0.18</td>
<td>0.8728</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nage Position</td>
<td>0.014957</td>
<td>0.006677</td>
<td>2.24</td>
<td>0.0306*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type: NagePosition</td>
<td>-0.013648</td>
<td>0.006677</td>
<td>-2.04</td>
<td>0.0498*</td>
</tr>
<tr>
<td>DE</td>
<td>Subject</td>
<td>(Intercept)</td>
<td>2.665889</td>
<td>0.025815</td>
<td>103.29</td>
<td>0.0001***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type</td>
<td>-0.002326</td>
<td>0.006436</td>
<td>-0.36</td>
<td>0.732</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nage Position</td>
<td>0.001038</td>
<td>0.006436</td>
<td>0.16</td>
<td>0.8666</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type: NagePosition</td>
<td>-0.010696</td>
<td>0.006436</td>
<td>-1.66</td>
<td>0.1042</td>
</tr>
<tr>
<td>MS</td>
<td>Subject</td>
<td>(Intercept)</td>
<td>2.710626</td>
<td>0.03739</td>
<td>72.5</td>
<td>0.0001***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type</td>
<td>-0.001512</td>
<td>0.008025</td>
<td>-0.19</td>
<td>0.8538</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nage Position</td>
<td>0.002577</td>
<td>0.008025</td>
<td>0.32</td>
<td>0.7458</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraction Type: NagePosition</td>
<td>-0.015972</td>
<td>0.008025</td>
<td>-1.99</td>
<td>0.0554</td>
</tr>
</tbody>
</table>

Note: "***": p < .001; "**": p < .01; "*": p < .05; ".": p < 1

At the region of RV/RS, there was no significant effect of Extraction types (p > .05). However, there was a marginal effect of position of Nage (p = .068). Importantly, there was an interaction between Extraction Types and Position of Nage (p < .05), as in Figure 3. The result shows that Nage-SRCs were read faster than SRCs-Nage, and ORCs-Nage were read faster than Nage-ORCs.

Figure 3 An interaction between the factors of Extraction Types and Position of Nage at the region of RV/RS (p < .01)
At the region of RO/RV, no significant effect of the Extraction Types was found \((p > .05)\). However, a significant effect on Position of Nage \((p < .05)\) and interaction between the Extraction Types and Position of Nage were observed \((p < .05)\). In other words, Nage in pre-RCs was read more slowly than Nage in post-RCs. The result can be explained more clearly on the interaction effect, as in Figure 4. The difference between Nage-SRCs and SRCs-Nage is much less than the difference between Nage-ORCs and ORCs-Nage. Specifically, Nage-ORCs were read more slowly than ORCs-Nage.

![Figure 4](image)

**Figure 4** An interaction between the factors of Extraction Types and Position of Nage at the region of RO/RV \((p < .05)\)

At the region of DE, there was no main effect and interaction \((p > .05)\).

At the region of MS, there was a marginal interaction between the Extraction Types and Position of Nage \((p > .05)\), as in Figure 5. The results show that Nage-SRCs were read faster than SRCs-Nage, whereas ORCs-Nage was read faster than Nage-ORCs. In other words, Nage shows the preference for the pre-SRCs and post-ORCs.

![Figure 5](image)

**Figure 5** A marginal interaction between the factors of Extraction Types and Position of Nage \((p = .0554)\) at the region of MS.
2.3 Discussion

In this experiment, two main factors (i.e. Extraction Types and Position of Nage) were examined in self-paced reading task. Importantly, the factor of Position of Nage was first evaluated in real-time processing, even though the previous literature has made contributions with the analysis of corpus data or/ and production experiment (Wu 2009, Ming 2010, Sheng and Wu 2012). All suggest that the positions of Nage denote different functions in discourse. That is, Nage in pre-RCs provides an early cue for listeners to anticipate RC structures, whereas Nage in post-RCs is able to avoid structural ambiguity (Sheng and Wu 2012). However, we propose that there is the other effect — the canonicity of thematic order in SVO languages, the default syntactic and semantic linking in online parsing.

The following sections will be discussed from two measures in self-paced reading experiment: (i) offline measures for the accuracy of yes/no comprehension questions in section 2.3.1 and (ii) online measures for log-transformed reading times in section 2.3.2.

2.3.1 Offline measures

The results show that Nage-SRCs were comprehended more accurately than SRCs-Nage (92% vs. 86%); by contrast, ORCs-Nage were comprehended more accurately than Nage-ORCs (90% vs. 86%). However, there are no main effects and interactions on the accuracy rate. This result is consistent with the findings from the sentence-production experiment (Sheng and Wu 2012) that the speakers had preference to produce Nage-SRCs, as compared to SRCs-Nage, whereas they had preference to produce ORCs-Nage, as compared to Nage-ORCs. In addition, our results were also consistent with the corpus-based analyses (Wu 2009, Ming 2010) that the frequencies of Nage-SRCs outnumber the frequencies of SRCs-Nage, whereas the frequencies of ORCs-Nage outnumber the frequencies of Nage-ORCs. Though a variety of the methodologies were employed above, they brought the same finding of the asymmetric patterns for the position of Nage in Chinese RCs.

2.3.2 Online measure

For the online data, the results were consistent with our predictions that Nage-SRCs is easier to process than SRCs-Nage; and by contrast, ORCs-Nage is easier to process than Nage-ORCs in accordance with the measures of reading times. This difference can be more clearly observed from the significant and the marginal interaction between the factors of Extraction Types and Positions of Nage, as shown in Figure 3, 4 and 5. The results for real-time reading also support the other corpus-based studies (Wu 2009, Ming 2010) and production experiment (Sheng and Wu 2012).

Moreover, the finding is attributed to the effect of the canonicity of thematic order (Lin 2012). That is, the structural preference (i.e. AGENT-ACTION-PATIENT for the dominant SVO word order in Chinese) for the readers plays a significant role in online parsing. Both Nage-SRCs and ORCs-Nage conform to the canonical thematic order in Chinese, whereas SRCs-Nage and Nage-ORCs do not. Thus, we suggest that
SRCs-Nage and Nage-ORCs may create additional processing cost due to a delayed thematic role assignment for their head noun. The examples illustrate in (12) and (13).

(12)a. **Nage-SRCs with canonical thematic order:**

```
[上午 __ 幫助  村民 的] 農夫 間候了  村長
AGENT ACTION PATIENT AGENT ACTION PATIENT
DCL ADV RV RO DE MS MV MO
```

Nage shangwu bangzhu cumin de nongfu wenhoule cunzhang
DCL morning help villager DE farmer greeted village head
“The farmer who helped the villager greeted the village head this morning.”

(12)b. **SRCs-Nage with non-canonical thematic order:**

```
[上午 __ 幫助  村民 的] 那個 農夫 間候了  村長
ACTION PATIENT AGENT ACTION PATIENT
ADV RV RO DE DCL MS MV MO
```

Shangwu bangzhu cumin de nage nongfu wenhoule cunzhang
morning help villager DE DCL farmer greeted village head
“The farmer who the villager helped greeted the village head this morning.”

(13)a. **Nage-ORCs with non-canonical thematic order:**

```
[上午  村民  幫助 __ 的] 農夫 間候了  村長
AGENT ACTION PATIENT
DCL ADV RS RV DE MS MV MO
```

Nage shangwu cumin bangzhu de nongfu wenhoule cunzhang
DCL morning villager help DE farmer greeted village head
“The farmer who the villager helped greeted the village head this morning.”

(13)b. **ORCs-Nage with canonical thematic order:**

```
[上午  村民  幫助 __ 的] 那個 農夫 間候了  村長
AGENT ACTION PATIENT ACTION PATIENT
ADV RS RV DE DCL MS MV MO
```

Shangwu cumin bangzhu de nage nongfu wenhoule cunzhang
morning villager help DE DCL farmer greeted village head
“The farmer who the villager helped greeted the village head this morning.”

In (12a), *Nage* in the initial position of the SRC can be assigned as a temporary *AGENT* so that the linear sequence follows the canonical thematic order. This *AGENT* role in the sentence-initial position can provide a cue for readers to anticipate an upcoming head noun. Thus, it facilitates the processing of head noun. However, *Nage*, which is adjacent to head noun *nongfu* in (12b), is not assigned as a temporary *AGENT*. In addition, when readers initially parse SRC with non-canonical thematic order (i.e. ACTION - PATIENT), they need to remain a track of *AGENT* for the later retrieval in the position of head noun. Thus, there is a delay in thematic role assignment when readers encounter the head noun which is assigned as an *AGENT*. 
Consequently, since (12a) follows the canonical thematic order in its linear sentence, the regions of RV and MS (or head noun) were read faster than those in (12b).

On the other hand, in (13a), Nage in the initial position of the ORC can be assigned as a temporary AGENT role, but readers realize that Nage is not an AGENT when they encounter RS cumin. Thus, readers need to remain track of Nage until they encounter MS (or head noun) nongfu that is assigned a delayed thematic role. By contrast, Nage, which is adjacent to head noun nongfu in (13b), is not assigned as a temporary AGENT. Readers initially parse the ORC with canonical thematic order (i.e. AGENT-ACTION), so the head noun is directly assigned as PATIENT. Consequently, since (15b) follows the canonical thematic order in ORCs, the regions of RS, RV and MS (or head noun) were read faster than those in (13a).

To sum up, this experiment tested two factors: (i) Position of Nage and (ii) Extraction Types. The results are compatible with the previous findings of corpus-based studies and the production experiment (Sheng and Wu 2012). The results are attributed to the effect of canonicity of thematic order (i.e. AGENT-ACTION-PATIENT in Chinese). If relative clauses conform to the canonical thematic order, they facilitate online processing for the regions of RC and head noun, and vice versa. In addition to frequency effect in corpus, we provide another important evidence for default semantic/syntactic linking in the online parsing. Lastly, we propose that Position of Nage indeed affects RC processing, and researchers should take this factor into account when they determine to use Nage into their experimental materials.

3. Conclusion

In this study, the position of Nage was examined to observe whether there is difference of RC processing for Nage in pre- vs. post-RC. Based on the previous corpus-based studies and production experiments, we expect that Nage-SRCs is easier to process than SRCs-Nage, whereas ORCs-Nage is easier to process than Nage-ORCs in accordance with online reading times. We evaluated four regions: the regions of RC (i.e. RV/RS and RO/RV) and the regions of matrix clauses (i.e. DE and MS). Our result was in line with the previous analyses (Sheng and Wu 2012). In addition to Sheng and Wu's (2012) discourse accounts, our findings also can be interpreted within the effect of canonicity of thematic order (Small et al. 2000, Lin 2012). The structures of Nage-SRCs and ORCs-Nage conform to the canonical thematic order like AGENT-ACTION-PATIENT, and hence this canonical order facilitates the processing at the regions of RC and head noun. By contrast, the structures of SRCs-Nage and Nage-ORCs do not conform to the canonical thematic order, and hence this non-canonical order interferes the processing at the regions of RC and head noun. The findings also suggest that there is a delayed thematic role assignment, and the extra processing cost in SRCs-Nage and Nage-ORCs during sentence processing. Importantly, the main processing difference for Nage position was interpreted as a processing demand for readers, and does not rely on the pragmatic functions only.
References


http://CRAN.R-project.org/package=languageR

Bates, Douglas, Martin Maechler and Ben Bolker. 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42.

http://CRAN.R-project.org/package=lme4


Tang, Ting-chi. 1979. 〈中文的關係子句〉，《國語語法研究論集》，249-300. 台灣：學生書局。

