The Effect of Classifiers in Predicting Chinese Relative Clauses
Fuyun Wu, Elsi Kaiser, & Elaine Andersen
University of Southern California

1. Introduction

Chinese is a typologically unique language with a combination of SVO word order and the N-final property (Dryer, 1992). It is also a classifier language in that each noun can be modified by a particular classifier depending on the semantic congruity between them. For instance, the classifier tiao can only modify a long, thin, and non-human entity. Thus it can modify a noun such as shengzi ‘string’, but not a human referent such as laoshi ‘teacher’.

In Chinese relative clauses (henceforth RC), there is a relativizer DE that occurs at the end of the RC, immediately preceding the head noun. This means that a comprehender may not identify an RC structure until very late. For example, a temporary ambiguity may arise in object-gapped RCs (as in (1)), because until encountering DE, the comprehender may initially interpret the initial noun and verb as constituting a matrix subject and matrix verb:

(1) [RC laoshi tuijian ti de ] shui
    teacher recommend t DE book

However, if the RC is preceded by the head noun’s classifier (e.g., the classifier ben for the upcoming noun shu ‘book’, as in (2a) below, and if this classifier is semantically incongruent with the immediately following noun laoshi ‘teacher’, then the presence of such a mismatching pre-RC classifier can provide a cue for the upcoming RC, which may facilitate the real-time parsing of head-final RCs. One way to test this is to compare reading times at the head noun region in a classifier mismatch condition (2a) to reading times at the same region in a classifier match condition (2b), where the classifier wei can modify a human referent (laoshi ‘teacher’) but not an inanimate entity (shu ‘book’). Faster reading times in the classifier mismatch condition would indicate facilitatory effects.
Reading time studies on the use of classifier mismatch as a cue in Chinese RC construction have not found a facilitatory effect in isolated sentences (Hsu, Phillips & Yoshida, 2005; Hsu, 2006). Instead, these studies found that mismatch conditions exhibited a slowdown at the local noun region (i.e., laoshi ‘teacher’). The classifier mismatch (ben in Ex. 2a), with its long-distance dependency relationship with the head noun (‘shu’), appears to disrupt lexical access to the immediately adjacent RC-internal noun. In contrast, classifier mismatch was found to facilitate RC processing when the target stimuli were presented in supportive discourse contexts, i.e., contexts where the existence of two previously-mentioned referents (e.g. two books) created the need to distinguish between them by means of an RC (Wu, Haskell & Andersen, 2006; Hsu, Hurewitz & Phillips, 2006). However, it is not clear how much of this RC-facilitation effect is due to discourse contexts which may have biased comprehenders towards an RC reading, and how much of the effect is due to classifier mismatch (Wu et al., 2006; Hsu et al., 2006).

All prior studies used classifiers that preceded the RCs (pre-RC classifiers). However, classifiers that modify the head noun can also occur after the RC, i.e., in a post-RC position, immediately before the head noun. Note that in order to provide a potentially useful cue for signaling an upcoming RC, the head noun’s classifier must occur in the pre-RC position, i.e., in a position that is dislocated from the noun it modifies. To find out how frequent such dislocated structures are in naturally-occurring language, we carried out a corpus study looking at the distribution of classifiers in pre-RC versus post-RC position. Based on these results, we designed two self-paced reading experiments to better investigate whether classifier positioning affects RC processing.

2. Corpus Study of Classifier Positioning

A total of 392 RCs that contain transitive action verbs subcategorizing two NP arguments were extracted from the Chinese Treebank 5.0 corpus. The RCs were then coded for 1) presence or absence of classifiers; and 2) the position of the head noun’s classifier (pre-RC or post-RC). We found that most Chinese sentences with RCs do not contain classifiers. This finding seems to suggest that classifiers as a functional category have a relatively special status in Chinese. This is perhaps not surprising since the most commonly used classifier in Chinese is the generic ge, which can refer to both a human referent and a
nonhuman entity, and is therefore usually dropped because it does not contribute much to the semantic content of the overall sentence meaning. Classifiers are, however, necessitated by the presence of a numeral or a demonstrative. Thus the use of classifiers is subject to special discourse contexts (such as previous mention) or referential events that specify number. Nevertheless, given the strict semantic congruence between a classifier and the noun it modifies, the presence of a classifier can potentially provide useful information for the language processing system.

Relevant to the issue of classifier positioning, Figure 1 illustrates the number of RCs of different syntactic types in which the head noun’s classifier occurred in pre-RC or in post-RC position in the corpus.

As shown in Figure 1, there is a mirror asymmetry for object-gapped and subject-gapped RCs: For object-gapped RCs (subject-modifying object-gapped, i.e., SO and object-modifying object-gapped, i.e., OO), classifiers appear mostly in post-RC position (84.85%, or 28/33), with only 5 instances (15.15%, or 5/33) in pre-RC position. Interestingly, this asymmetry is reversed for the subject-gapped RCs (subject-modifying subject-gapped, i.e., SS and object-modifying subject-gapped, i.e., OS). For OS RCs, pre-RC classifiers occur more frequently than post-RC classifiers: out of the 24 OS RCs with classifiers, 21 (87.5%) have pre-RC classifiers, and only 3 (12.5%) have post-RC classifiers. For SS RCs, the asymmetry between pre-RC and post-RC classifiers is less dramatic: Out of the 38 SS RCs with classifiers, 23 (60.53%) have pre-RC classifiers and 15 (39.47%) have post-RC classifiers.
We posit two processing-driven principles to account for the asymmetrical distribution of classifiers: (i) the **Early Occurrence Strategy** and (ii) the **Semantic Clash Avoidance Principle**. The Early Occurrence Strategy states that classifiers prefer to occur as early as possible. This strategy fits with the growing body of research on anticipatory and expectation-based processing (e.g., Altmann and Kamide, 2004; Levy, 2007). This strategy can explain the pre-RC preference for subject-gapped RCs. But what about the post-RC preference that we see for object-gapped RCs? The Semantic Clash Avoidance Principle states that classifier positioning should avoid disrupting lexical access to the following noun. Since classifiers are adjacent to the ‘wrong’ noun (i.e., the embedded RC-subject) in object-gapped RCs, pre-RC classifiers are therefore dispreferred in this context. That is, object-gapped RCs prefer post-RC classifiers. To further test these two principles, we conducted two self-paced reading experiments.

3. Experiment 1

Experiment 1 was designed to test the Early Occurrence Strategy. We hypothesized that if pre-RC classifiers have a facilitatory effect on RC structure-building, their presence should result in faster reading times at the head noun region than when there is no classifier. Furthermore, given that pre-RC classifiers are adjacent to the wrong noun in object-gapped RCs, which violates the Semantic Clash Avoidance Principle, our expectation was that pre-RC classifiers should be more helpful predictors in subject-gapped RCs than in object-gapped RCs. We therefore predicted that when pre-RC classifiers are present, we should see faster reading times at the head noun region in subject-gapped RCs than in object-gapped RCs.

Forty native speakers of Mandarin Chinese participated. We manipulated extraction type (subject- vs. object-gapped) and classifier (absent vs. present), with stimuli consisting of 24 sets of targets in four conditions and 44 fillers. An example set of the four conditions is given in (3).

(3) a. **Subject-gapped RCs without classifiers**

\[
\text{e, zaizhong jizhe \_ de jushi, mimide zhangzhe qingtai} \\
\text{hit journalist DE boulder thickly grow moss} \\
\text{‘The boulder that \_ hit the journalist is thickly covered with moss’}. \\
\]

b. **Subject-gapped RCs with classifiers**

\[
\text{na-kuai e, zaizhong jizhe \_ de jushi, mimide zhangzhe qingtai} \\
\text{that-CLboulder hit \_ journalist DE boulder thickly grow moss} \\
\text{‘The boulder that \_ hit the journalist is thickly covered with moss’}. \\
\]

c. **Object-gapped RCs without classifiers**
jushi zaizhong ei de jizhei        jingtide     huangu       sizhou.
‘The journalist that the boulder hit __ looked about his surroundings cautiously.’

d. **Object-gapped RCs with classifiers**
na-wei jushi zaizhong ei de jizhe, jingtide huangu sizhou.
that-CL human boulder hit DE journalist cautiously look-about surroundings
‘The journalist that the boulder hit __ looked about his surroundings cautiously.’

3.1 Results and discussion

Since the classifier-absent conditions in two types of RCs did not contain the sentence-initial Dem+CL sequence, their first word positions were aligned with the second word positions in the classifier-present conditions. The analyses focused on the seven words following the Dem+CL sequence (pos1): The RC-internal verb (V) or the RC-internal subject (S, pos2), the RC-internal object (O) or the RC-internal verb (V, pos3), the relativizer DE (pos4), the head of the RC (S or O, pos5), the post-head adverb (ADV, pos6), the matrix verb (MV, pos7), and the postverbal matrix object (MVO, pos8). Figure 2 presents mean reading times for these eight positions in the target stimuli for the four conditions.

![Graph showing mean reading times for different conditions](image-url)
As shown in Figure 2, at word position 2 (i.e., the RC-internal object jizhe ‘journalist’ in subject-gapped RCs or the RC-verb zaizhong ‘hit’ in object-gapped RCs), there was a classifier disruption effect in object-gapped RCs, but not in subject-gapped RCs: SO RCs with classifiers were read significantly slower than SO RCs without classifiers; there was no difference between the classifier-present and the classifier-absent conditions in subject-gapped RCs.

At the Adverb (position 6), there was a classifier facilitation effect in subject-gapped RCs, but not in object-gapped RCs: Reading times in SS RCs without classifiers were significantly slower than SS RCs with classifiers. In addition, object-gapped RCs were read significantly slower than subject-gapped RCs.

At the main verb (position 7), the facilitatory effect of classifier was found in both subject-gapped and object-gapped RCs: Reading times in classifier-absent conditions were slower than reading times in classifier-present conditions.

At the main object (position 8), SS RCs without classifiers were slower than SS RCs with classifiers.

Thus the results showed classifier disruption effects in object-gapped RCs at the embedded noun region, replicating the lexical disruption effects found in prior work and predicted by the Semantic Clash Avoidance Principle. But later in the sentence (from the adverb region to the sentence-final region), we found facilitatory effects of classifier for both subject-gapped RCs and object-gapped RCs. This supports the Early Occurrence Strategy. Also, the reading time differences in subject-gapped RCs were numerically bigger than those in object-gapped RCs, indicating that the facilitatory effects of pre-RC classifiers as RC-predictors are greater in subject-gapped RCs than in object-gapped RCs.

In sum, the outcome of Experiment 1 showed that, as predicted by the Early Occurrence Strategy, pre-RC classifier-present conditions were read faster than classifier-absent conditions.

4. Experiment 2

The results of Experiment 1 suggest that the facilitatory effect of pre-RC classifiers may be weaker for object-gapped RCs than for subject-gapped RCs. This may be due to the lexical disruption effect incurred early in object-gapped RCs; that is, the classifier mismatch caused such great difficulty in accessing the following noun that it adversely affected the overall parsing processes of RC structure building. Experiment 2 was conducted to further test potential effects of pre-RC classifiers as cues in object-gapped RCs. Thus, this experiment investigates issues related to the Semantic Clash Avoidance Principle, focusing solely on object-gapped RCs. A control condition was needed to facilitate the
comparison between the classifier-present and classifier-absent conditions. The passive BEI sentences were chosen to fulfill this purpose.

BEI in Chinese is a passive morpheme. In passive constructions with this marker, BEI is preceded by a noun that denotes the patient and may be followed by an optional noun that denotes the agent. For instance, in the object-gapped RC fragment *nashan zuqiu dasui de huaping* ‘that-CL football break DE window’ (the window that the football broke), the passive marker BEI can be added in front of the embedded RC-subject *zuqiu* ‘football’ to clearly mark it as the agent of the breaking event. Crucially, BEI intervenes between the classifier *shan* and the local noun *zuqiu* ‘football’, i.e., right where the semantic clash due to classifier mismatch would otherwise occur. Furthermore, since the noun following BEI (*zuqui* football) naturally forms a constituent indicative of the agent of a certain event (i.e., ‘break’) that is suggestive to an anticipatory parser of a recipient or a patient that undergoes the event, there is no other way to continue the sentence except with an RC. Thus, with BEI, we can create sentences that are clearly RCs but have the same word order as RCs without BEI.

In Experiment 2, we use the BEI construction as a control condition, and compare the processing consequences of BEI with those of classifier mismatch. According to the Semantic Clash Avoidance Principle, a pre-RC classifier mismatch anchored at the left periphery may cause a semantic clash in object-gapped RCs. However, we hypothesize that the presence of BEI may render semantic clash effects vacuous, because BEI provides clear evidence of an RC and thus indicates that the subsequent noun is not associated with the pre-RC classifier. We therefore predict that pre-RC classifiers are a weaker cue for RC structure building than the syntactic cue provided by BEI in object-gapped RCs.

We tested 32 native speakers of Chinese. The critical items crossed classifier (presence vs. absence) and BEI (presence vs. absence), yielding four conditions (as in (3)). Experiment 2 contained 24 critical items and 48 filler items.

(3) a. no-CL, no-BEI

jushi zaizhong e, de jizhe, jingtide huangu sizhou.

‘The journalist that the boulder hit _ cautiously looked about his surroundings.’

b. CL, no-BEI

na-wei jushi zaizhong e, de jizhe, jingtide huangu sizhou.

‘The journalist that the boulder hit cautiously looked about his surroundings.’

c. no-CL, BEI

ei bei jushi zaizhong de jizhe, jingtide huangu sizhou

BEI boulder hit DE journalist cautiously look-about surroundings
‘The journalist that _ was hit by the boulder cautiously looked about his surroundings.’

d. **CL, BEI**
na-wei e, bei jushi zaizhong de jizhe, jingtide huangu sizhou that-CLhuman BEI boulder hit DE journalist cautiously look-about surroundings
‘The journalist that _ was hit by the boulder cautiously looked about his surroundings.’

**4.1 Results and discussion**

Since the four conditions differed in whether the classifier and BEI were present or absent, word positions were aligned by syntactic categories across conditions, resulting in the common word beginning at the third position, the RC-internal subject. The analyses focused on the seven words following the Dem+CL sequence (pos1) and BEI (pos2): the RC-internal subject (S, pos3), the RC-internal verb (V, pos4), the relativizer DE (pos5), the head of the RC (O, pos6), the post-head adverb (ADV, pos7), the matrix verb (MV, pos8), and the postverbal matrix object (MVO, pos9). Figure 3 presents mean reading times for each of the seven word positions of the target stimuli in the four conditions.

Figure 3: Mean reading times per word position in Experiment 2. RTs are in milliseconds.
As shown in Figure 3, at both the RC-verb region (pos 4) and the relativizer DE region (pos 5), reading times in the classifier, no-BEI condition were significantly slower than the other three conditions, which fits with the idea that classifier mismatch results in processing difficulty due to semantic clash.

At the RC head-noun region (pos 6), reading times in the BEI-absent conditions were significantly slower than reading times in the BEI-present conditions, suggesting a facilitatory effect of BEI.

At the adverb region (pos 7), there was again a facilitatory effect of classifier: Reading times in the no-classifier, no-BEI condition were slower than reading times in the classifier, no-BEI condition. At the adverb region, there was also a facilitatory effect of BEI. In fact, no reading time differences were found between the classifier, BEI condition, the classifier, no-BEI condition, and the no-classifier, BEI condition, presumably because the presence of a cue—be it a classifier or a BEI, facilitates processing towards the end of sentence.

These facilitatory effects of classifier and BEI continued throughout the main verb region (pos 8) and the main object region (pos 9).

In sum, Experiment 2 replicated the classifier facilitatory effect. The classifier-absent, BEI-absent condition was read slowest. Reading times also provide further evidence for the Semantic Clash Avoidance principle, as the classifier-present, BEI-absent condition was slowest when followed by the incongruent embedded noun (at positions 4 and 5). We also found evidence that BEI is a more effective cue for an upcoming RC than classifier mismatch; Bei seems to prevent any semantic clash between the mismatching classifier and the local noun: Sentences with BEI were read faster overall regardless of whether a classifier was present; when both BEI and a classifier were present, the sentences were read fastest.

5. Conclusions

In this paper, we reported the results of a corpus study and two experiments on the effect of classifiers on RC processing in Chinese. The corpus analysis revealed an asymmetrical distribution pattern for classifiers in RCs of different extraction types. Two hypotheses were formulated to account for this asymmetry: the Early Occurrence Strategy and the Semantic Clash Avoidance Principle. The behavioral data from the two self-paced reading experiments support these two hypotheses. Overall, we found that processing of RCs in Chinese is facilitated if there is a classifier cue; and that pre-RC classifiers are more helpful in subject-gapped RCs than in object-gapped RCs. This study provides evidence for classifier mismatch serving as a RC-predictor, even in the absence of preceding discourse context.
References


Fuyun Wu, Elsi Kaiser, & Elaine Andersen
University of Southern California
Department of Linguistics
Los Angeles, CA 90089-1693
fuyunwu@usc.edu
elsi.kaiser@usc.edu
eanderse@usc.edu