

## **Chinese and English Relative Clauses: Processing Constraints and Typological Consequences**

Chien-Jer Charles Lin  
*Indiana University*

Languages with distinct typological characteristics often present different challenges to language processing and thus lead to differences in linguistic performances across languages. This study investigated the relations between linguistic typology, constraints on language processing, and crosslinguistic differences in language use. Two studies on Chinese and English relative clauses were conducted. The first study looked at the semantic properties of the head nouns in Chinese and English relative clauses, showing that shared communicative functions lead to shared patterns in Chinese and English relative clauses: the heads of object relative clauses tend to be inanimate nouns, while those of subject relative clauses tend to be animate. The second study looked into the use of relative clauses in parallel texts, showing that the typological differences in terms of head positions in Chinese and English led to distinctive complexity patterns. Chinese relative clauses tend to be shorter (in terms of number of syllables and number of words) and less complex (in terms of embeddings other relative clauses) than English relative clauses. These results supported the Head-driven Constituent Complexity Hypothesis.

### **1. Introduction**

Languages with distinct typological characteristics often present different challenges to language processing. Such differences lead to differences in linguistic performances across languages. Previous attempts to associate language processing with linguistic typology have focused on two themes: using universal processing constraints to account for crosslinguistic differences and finding functional contrasts in different word orders. The first approach was exemplified by the seminal typological study of relative clauses conducted by Keenan and Comrie (1977), who investigated the relativizability of noun phrases at different grammatical positions across languages and found that different grammatical positions showed differences on the ease of extraction across languages. Noun phrases in certain grammatical positions are more likely to be relativized than noun phrases in other positions. They proposed the well-known Keenan-Comrie Accessibility Hierarchy (i.e, Subject > Direct Object > Indirect Object > Oblique object > Genitive > Object Complement, abbreviated as the AH), which summarized the likelihood of relativization across languages. When a language can relativize noun phrases at a gram-

matical position, it can also relativize noun phrases at all higher positions on the hierarchy. For this generalization, they provided a processing account: noun phrases higher in the hierarchy are easier to access than those that are lower in the hierarchy. They explicitly attributed this accessibility hierarchy to “the psychological ease of comprehension”, namely, that “the lower a position is on the AH, the harder it is to understand relative clauses formed on that position (p.88).”

More recently, Langus and Naspor (2010) entertained the hypothesis that different cognitive functions motivate two popular word orders in human languages—SVO and SOV, which together accounted for 76% of human languages (Dryer, 2005). According to this hypothesis, SVO is a preferred word order based on the computational system of human syntax, while SOV is a preferred form for effective communication. The support for SOV being communicatively motivated came from their experiments asking Italian and Turkish speakers to describe events by using gestures, not the spoken language. They found that despite the different basic word orders in Italian (SVO) and Turkish (SOV), SOV was the preferred order in gesturing. In a comprehension study of different word orders using gesture inputs, they also found the SOV order to take the least time for comprehension. Interestingly, however, in an experiment where participants listened to words and sentences in their native languages presented with prosodically flat speech, regardless of the dominant word orders in their native languages, VO is preferred over OV (when the subject position was controlled for). The preference of VO over OV is taken to support the computational superiority of SVO over SOV.

The current research provides a different take on the relation between processing and crosslinguistic contrasts. Our study examined two processing-based hypotheses by looking at relative clauses in large corpora as well as those in parallel translated texts. The corpus study looked at the animacy of the head nouns in Mandarin Chinese and compared the findings with those of the processing studies of other languages. The study of parallel translated texts examined the complexity of relative clauses in head-initial (English) and head-final (Chinese) structures.

Before we delve into the data, a few notes about Chinese and English relative clauses are in order. Chinese and English are both languages that rely heavily on word orders for coding thematic relations. The dominant word orders for both languages are SVO, semantically interpreted as Agent-Action-Patient. In terms of head positions of nominal structures, however, the two languages contrast typologically. Chinese noun phrases are head-final; all modifying phrases, including adjectivals, prepositional adverbials, and relative clauses, precede the nouns. English noun phrases are head-initial; the preposition phrases and relative clauses follow rather than precede the head nouns. Such a contrast makes Chinese and English an ideal pair of languages for investigating the effect of head positions in processing as well as language use. They are typologically distinctive in terms of the head positions in noun phrases; yet, they are typologically similar in terms of the basic word orders. Such a contrast allows us to attribute differ-

ences in processing and language use to the position of the heads internal to noun phrases rather than to other factors.

In section 2 of this article, we report a corpus study of Chinese relative clauses, looking at the animacy of the head nouns. Section 3 investigated how the typological differences in terms of head positions in Chinese and English led to distinctive clausal complexity of Chinese and English relative clauses in parallel translated texts. Section 4 concludes the paper.

## **2. Study 1: Animacy of the head in Chinese and English relative clauses**

Several recent processing studies of relative clauses have shown that the animacy of the head noun affects how easy relative clauses are for comprehension. Mak, Vonk, and Schriefers (2002) demonstrated that while subject relatives are generally easier to process than object relatives in Dutch, when the head nouns of the object relatives were inanimate, there were no processing differences between subject and object relatives. Mak, Vonk, Schriefers (2006) further demonstrated that animacy on both the head noun and the noun phrases inside the relative clauses is important in processing.

Gennari and MacDonald (2008) also investigated the effect of animacy on English relative clauses. They conducted a gated sentence completion task using animate and inanimate nouns as the head nouns. Only 15% of the sentences with animate head nouns were completed as object relative clauses, while 65% of the sentences with inanimate head nouns were completed as object relatives. These results were corroborated by reading time data.

Regarding the animacy of the head nouns of Chinese relative clauses, Wu, Kaiser, and Anderson (2009) extracted 331 relative clauses from the Chinese Treebank 5.0 corpus (Palmer, Chiou, Xue & Xia, 2005). They examined the animacy information on the head noun and the embedded noun in subject and object modifying relative clauses that involved subject and object extractions and found that subject extracted relative clauses tended to have animate heads and inanimate embedded nouns, while object extracted relative clauses tended to have inanimate heads and animate embedded nouns. The self-paced reading experiments they conducted showed the same preference for contrastive animacy on the head noun and the embedded noun. Wu, Kaiser, and Anderson (2011) further extracted 1218 relative clauses in a later study. They found that the heads of object relatives are overwhelmingly inanimate, but the animacy of the heads of subject relatives depended on whether the relative clauses modified the subject or the object of the main clause. If the relative clause modifies the subject, then its head noun tends to be animate. If it modifies the object, then its head noun is equally likely to be animate or inanimate.

The current study is interested in whether the animacy information on the head nouns of Chinese relative clauses would follow the same universal pattern; that is, whether we would find more inanimate head nouns in object relatives and more animate head nouns in subject relatives. In comparison with previous studies, our study collected

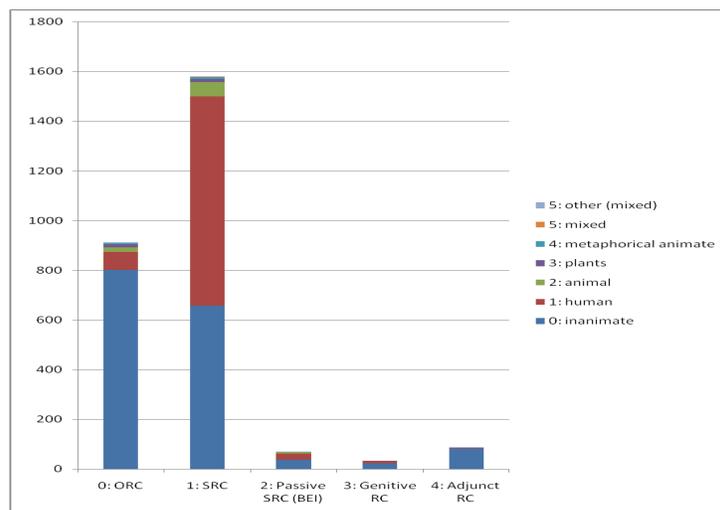
a greater number of relative clauses and classified these relative clauses into finer categories. 3075 relative clauses were extracted from Sinica Treebank 3.0 released by Academia Sinica in Taiwan (<http://godel.iis.sinica.edu.tw/CKIP/treebank.htm>). These relative clauses were manually coded on syntactic and semantic dimensions. In previous studies, the passive relatives, the possessive relatives, and the adjunct relatives were not coded separately. Our study makes it possible to look at animacy of these relative clauses as well. For the purpose of the current study, we categorized relative clauses into the following types for illustration:

(1) Classification of relative clauses extracted from Treebank 3.0:

Type	Example
0: ORC	張老師處罰的學生
1: SRC	沒有考過的人
2: Passive SRC	被處罰的學生
3: Possessive RC	家被颱風吹垮的居民
4: Adjunct RC	張老師處罰學生的原因

In terms of the overall distribution, more subject relatives were found than object relatives, confirming previous findings on the dominant frequencies of subject relatives over object relatives. All other types of relative clauses were relatively infrequent.

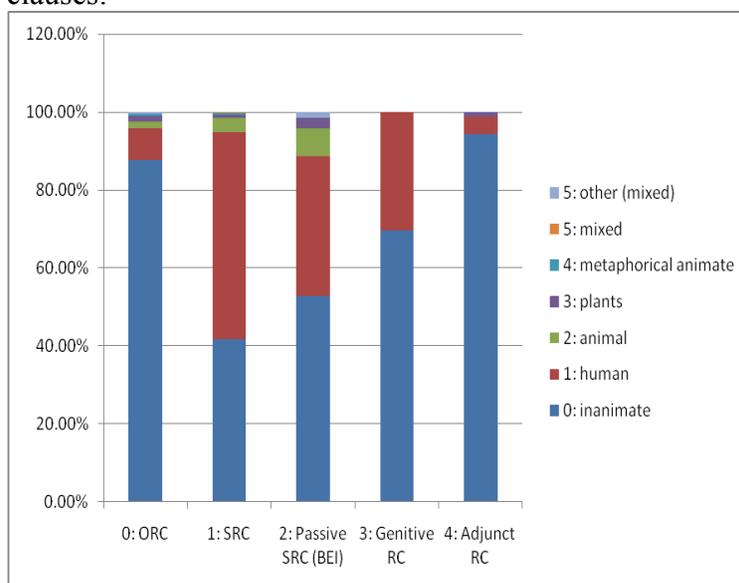
(2) Raw frequencies of different kinds of relative clauses and animacy of the head nouns:



Semantically, the head nouns were coded as inanimate, human, animal, plant, and metaphorically animate. Within each type of relative clauses, different distributions regarding head noun animacy were found. Close to 90% of object relatives in Chinese

have inanimate head nouns, while animate and inanimate head nouns were more equally distributed on subject relatives (though there were more animate head nouns than inanimate head nouns). The dominance of inanimate head nouns on object relatives is consistent with a similar dominance found in English and Dutch. These results reflect a universal tendency for the patient roles (located at the object positions) to be inanimate. A similar preference was found on passive relatives; even though the extracted head nouns of passive relative clauses are at the subject position in the embedded clauses, they usually hold the patient role thematically. Therefore, these head nouns are more prone to be inanimate and nonhuman.<sup>1</sup>

(3) Distribution of animacy information on the head nouns of different kinds of relative clauses:



These findings confirmed a universal preference for inanimate head nouns in object relative clauses. The head nouns of subject relatives, however, are less biased in terms of animacy (similar to the equivalence of animate and inanimate head nouns in object modifying subject-extracted relatives obtained by Wu et al. 2011).

### 3. Study 2: Chinese and English relative clauses in parallel (translated) texts

The second study focused on how the differences of head positions may result in different degrees of complexity in the production of head-initial and head-final relative clauses. We propose the *head-driven constituent complexity hypothesis* (HCCH) based on the observations that typologically opposite head positions may induce distinct pro-

<sup>1</sup> Note, however, that the subjects of passives in Mandarin Chinese are not overwhelmingly inanimate. Close to 50% of these head nouns were actually animate nouns.

cessing strategies (Lin, 2011) and that the later the head noun is encountered, the greater temporary uncertainty exists in parsing, and therefore the more difficult for parsing. We therefore hypothesize that head-final structures are overall harder than head-initial structures due to the uncertainty prior to the appearance of the heads. The HCCH predicts that given the same contexts and similar textual contents, head-final constituents tend to be shorter and less complex than head-initial constituents.

To test this hypothesis, we compared the complexity of relative clauses in English and Chinese texts. This paper presents only the preliminary results based on our textual analysis of two sets of comparable essays. However, the contrasts were distinctive enough and therefore the results were quite suggestive even though the data were limited at the current phase. The complexity of relative clauses was measured by the length of the relative clauses and the structural complexity of embeddings. The lengths of relative clauses were measured by (a) the number of syllables before the head noun in an embedded clause and (b) the number of words (i.e., semantic units) before the head noun in an embedded clause. An embedded clause is structurally complex when it has another relative clause embedded in it.

Two comparable popular magazines were selected as the targeted texts--*Taiwan Panorama* (*Taiwan Guanghai Zazhi* 台灣光華雜誌) and *Scientific American* (*Kexueren* 科學人). An article from the *Taiwan Panorama*, which was originally written in Chinese and translated into English, was randomly selected for analysis. This article was made of 128 English sentences.<sup>2</sup> In the corresponding Chinese text, 82 relative clauses were identified.<sup>3</sup> The average lengths of the Chinese relative clauses were 6.63 in terms of number of syllables and 3.78 in terms of number of words. In the translated English text, 48 relative clauses were identified, and the average length of these relative clauses was 11.28 syllables long corresponding to 6.83 words. Therefore, both in terms of phonological length and semantic complexity, head-initial relative clauses tend to be longer than head-final relative clauses. Among the Chinese relative clauses, no relative clauses were embedded in other relative clauses. Among the English relative clauses, 4 of them were embedded in other relative clauses. English relative clauses are more tolerant of structural complexity than Chinese relative clauses.

To be sure that the difference was not due to translated texts versus original texts, we also analyzed relative clauses in an article originally written in English (published in *Scientific American*) and its Chinese translation published in the magazine *Kexueren* (科學人). In the original English text, which contained 111 sentences, 39 relative clauses were identified. The average number of syllables in the English relative clauses was

---

<sup>2</sup> Number of sentences were measured based on the English texts because the definition of what a sentence is more definite in English than in Chinese.

<sup>3</sup> Relative clauses in Chinese are defined as clauses embedded in DE-phrases with verbs (including stative verbs) in them. English relative clause are defined as clauses embedded in noun phrases with relativizers such as, *who*, *which*, *that*, *whom*, *whose*, *what*, *where*, *why* in them, and in noun phrases with embedded verbs in the forms of past participles and gerunds.

13.47, which included an average of 7.39 number of words. 77 relative clauses were identified in the Chinese text. The average length of the Chinese relative clauses was 7.32 syllables and 4.51 words long. Among the relative clauses, 3 relative clauses in English were embedded in relative clauses, and no relative clauses in Chinese were embedded in other relative clauses.<sup>4</sup>

- (4) Lengths of Chinese and English relative clauses in the parallel texts (number of syllables):

	Chinese to English	English to Chinese
Chinese relative clauses	6.63	7.32
English relative clauses	11.28	13.47

- (5) Lengths of Chinese and English relative clauses in the parallel texts (number of words):

	Chinese to English	English to Chinese
Chinese relative clauses	3.78	4.51
English relative clauses	6.83	7.39

- (6) Number of relative clauses embedded in other relative clauses in the Chinese-English parallel texts:

	Chinese to English	English to Chinese
Chinese relative clauses	0	0
English relative clauses	4	3

These results confirmed the HCCH. Regardless of the direction of translation, Chinese relative clauses are on average shorter than English relative clauses both in terms of phonological length (i.e., number of syllables) and in terms of semantic units (i.e., number of words). English relative clauses are more likely to have relative clauses embedded in them but no occurrences of multiply embedded relative clauses have been found in Chinese. Both in terms of semantic and phonological quantity and in terms of structural complexity, therefore, head-initial relative clauses tend to be more complex than head-final relative clauses.

#### 4. Concluding remarks

In this paper, we investigated two aspects of Chinese and English relative clauses—animacy of the head nouns and clausal complexity. Previous studies of head-initial relative clauses (in Dutch and in English) showed that animacy of the head is important in determining whether an object relative clause is easy or difficult to comprehend. An

<sup>4</sup> Note that 3 Chinese relative clauses were embedded in genitive DE-phrases, which, though not considered as relative clauses in our study, appeared to be structurally complex nevertheless.

object relative clause with an inanimate head noun and an animate embedded noun was no more difficult than a subject relative clause. Given shared functions of relative clauses across languages (i.e., modifying nouns and providing grounding information in a sentence), it is expected that animacy patterns on the head would be shared across languages. Our study of the head nouns in Chinese relative clauses in corpora indeed found overwhelming use of inanimate nouns as the head nouns of object relative clauses in Chinese. The universal pattern suggests that human languages tend to use animate nouns as the subject and inanimate nouns as the object.

The second study looked into the complexity of Chinese and English relative clauses. We tested the Head-Driven Constituent Complexity Hypothesis, which predicted that head-initial relative clauses are more tolerant of clausal complexity than head-final relative clauses. Given that Chinese and English shared the same word orders, our hypothesis is that any difference found between Chinese and English relative clauses may be due to the opposite positions of the heads inside noun phrases. Our comparisons of parallel texts of Chinese and English showed that in both texts translated from Chinese to English and those translated from English to Chinese, Chinese relative clauses were shorter and less tolerant of complexity than English relative clauses.

For future research, the corpus study on animacy should also consider the information status of these noun phrases and whether the relative clauses modified the subject or the object of the matrix clauses. The comparison between Chinese and English relative clauses can be better understood when the complexity of other aspects of Chinese and English structures are also contrasted.

### References

- Gennari, Silvia P. and Maryellan C. MacDonald. 2008. Semantic indeterminacy and relative clause comprehension. *Journal of Memory and Language* 58.161-87.
- Dryer, Matthew. 2005. The order of subject, object and verb. *The world atlas of language structures*, ed. by M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie, 330-333. Oxford: Oxford University Press.
- Keenan, Edward L. and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8.63-99.
- Langus, Alan and Marina Nespore. 2010. Cognitive systems struggling for word order. *Cognitive Psychology* 60.291-318.
- Lin, Chien-Jer C. 2011. Thematic templates and the processing asymmetry of Chinese relative clauses. Manuscript. Indiana University, Bloomington.
- Mak, W. M., W. Vonk and H. Schriefers. 2002. The influence of animacy on relative clause processing. *Journal of Memory and Language* 47.50-68.
- Mak, W. M., W. Vonk and H. Schriefers. 2006. Animacy in relative clauses: The hikers that rocks crush. *Journal of Memory and Language* 54.466-490.
- Palmer, M., F.-D. Chiou, N. Xue and F. Xia. 2005. The Penn Chinese Treebank Corpus. University of Pennsylvania.

- Wu, Fuyun, Elsi Kaiser and Elaine Andersen. 2009. Animacy effects in Chinese relative clause processing. *Proceedings of the 38th western conference on linguistics*, ed. by Michael Grosvald and Dionne Soares, 318-329. Davis: University of California, Davis.
- Wu, Fuyun, Elsi Kaiser and Elaine Andersen. 2011. Subject preference, head animacy, and lexical cues: A corpus study of relative clauses in Chinese. *Processing and producing head-final structures*, ed. by Hiroko Yamashita, Yuki Hirose, and Jerry Packard, 173-194. The Netherlands: Springer.