

**International Symposium on Diverse Approaches to
Second Language Acquisition: Learner Corpora,
Evaluation and Brain Sciences
June 1st-2nd 2019, TUFS, Tokyo, Japan**

**The Production of Relative Clauses
in L2 Chinese:
A Corpus-based Study**

Li-ping Chang

Graduate Program of Teaching Chinese as a Second Language
National Taiwan University

Background

- Many previous studies indicate that as learners acquire relative clause (RC), the difficulty order conforms to the Noun Phrase Accessibility Hierarchy (AH). That is, the subject-extracted RC (Subj-RC, as shown below) is the easiest to be acquired (Gass, 1979; Eckman, Bell & Nelson, 1988; Doughty, 1991; Izumi, 2003).
- Early studies are mostly focused on Indo-European languages and majorly English.

(1) the man [who bought the book]

(N-Rel)(English Subj-RC)

(2) [*mai shu de*] ren [买书的] 人

(Rel-N)(Chinese Subj-RC)

buy book NOM person

‘the person who bought the book’

Background

- According to the typological research (Dryer 1992), VO order implies N-Rel order whereas Rel-N order implies OV order as in example (3). However, Chinese is a VO language, but the order of relative clause and head noun is Rel-N rather than N-Rel. It is very idiosyncratic.

(3) *gakusei-ga* ____ *katta* *hon*
student-NOM bought book

Background

- The question of whether languages such as Japanese or Mandarin Chinese, which place the head noun at the right, also conform to the AH prediction has been the subject of inquiry for decades. There are dispute results occurring.
- For example, Tarallo & Myhill's (1983) cross-language research indicates that as English natives learning languages where the RC occurs after a head noun, including German, Portuguese, and Persian, the Subj-RC proves easier than the Obj-RC. This appears to conform to the AH prediction, but they also find the reverse occurs when English native speakers learning Japanese or Chinese, the Obj-RC is easier than the Subj-RC.

Research Questions

- Is Subj-RC or Obj-RC easier for L2 Chinese learners?
- For L2 Chinese learners, do different typological L1s show different tendency of RCs?

Previous Studies: *Dispute Results on the Acquisition of Subj-RC and Obj-RC*

- Packard's (2008) research utilizes the self-paced reading task to assess the L2 Chinese processing difficulty of RCs by English speakers. The result showed that English speakers demonstrate slower processing times for Subj-RCs. That means the Obj-RCs are easier.
- Different approach by Dai (2010) also supported that the Obj-RCs are easier than Subj-RCs. He employed the combining-sentence task. The 39 testees are from English, Japanese, and Korean speakers.
- It seems the acquisition of L2 Chinese RCs does not support the AH hypothesis.

Previous Studies: *Dispute Results on the Acquisition of Subj-RC and Obj-RC*

- There are other studies which suggest that the AH acquisition theory is applicable to L2 Chinese.
- For example, Xu (2014) conducted a combining-sentence task for 45 native speakers of English in order to investigate if the order of difficulty conforms to the AH prediction. The results showed that the learners preferred to produce Subj-RCs. And she also claimed that Subj-RCs were easier than Obj-RCs through the analysis of learner's response accuracy.

Previous Studies: *Effects of animacy of head nouns*

- Apart from the predictive power of the AH, analysis based on the animacy of head nouns has provided much insight into the study.
- For example, Traxler, Morris & Seely (2002) use eye-tracking testing to conclude that an Obj-RC followed an animate head noun is more difficult to process, such as “The mountaineer that the boulder hit”, than an inanimate head noun, such as “The rock which the boy threw.” This shows that the animacy of a head noun is related to the difficulty of comprehension of an RC.
- The results of Ozeki & Shirai (2007) also support the effect of animacy.

Previous Studies: *Effects of animacy of head nouns*

- Regarding L2 Chinese study, Li (2015) demonstrated that the animacy of nouns in the RCs strongly affects the production of RC type by observing the HSK corpus. He also declares that AH is second to animacy in affecting the production of RC type.
- However, his research did not take learners' proficiency into account and observed only limited amount of 201 samples. In our study, we also adopted the corpus approach and observed 2259 RCs from 6 various L1s and two proficiency levels. Hopefully a better profile can be derived to settle the dispute of the above inconsistent results.

Previous Studies: *Effects of positions in a matrix sentence for Subj-RCs and Obj-RCs*

- Some cognitive theories posit that center-embedded RCs may interrupt language processing, therefore, they are more difficult to comprehend than those (right- or left embedded) occur on the sides of the matrix sentence (Bever, 1970, Kuno, 1974).
- In such view, Chinese RC nominals at subject position (either SS or SO) should be easier to process than object-position RC nominals (either OS or OO) as shown in the examples below.

Either SS or SO should be easier to process.

- (4) *ta bu shi [na ge mai shu de ren]_{OS}*. 他不是[那個買書的人]_{OS}。
he not be that-Cl buy book de person
'He is not [the one who bought that book].'
- (5) *ta xihuan [Zhangshan mai de shu]_{OO}*. 他喜歡[張三買的書]_{OO}。
he likes Zhangsan buy de book
'He likes [the books which Zhangsan bought].'
- (6) [*mai shu de nage ren*]_{SS} *bu shi wo tongxue*. [買書的那個人]_{SS}不是我同學。
buy book de that-Cl person not be my classmate
'[The one who bought that book] is not my classmate.'
- (7) [*Zhangsan mai de shu*]_{SO} *bu jian le*. [張三買的書]_{SO}不見了。
Zhangsan buy de book not see asp
'[The books Zhangsan bought] are lost.'

Aim of the Study

- The goal of our research is to uncover the RC production of L2 Chinese learners in different aspects:
 - a) Subj-RCs , Obj-RCs and their position in a sentence,
 - b) different proficiency levels of learners and
 - c) learners' native languages.

Methodology

- Many previous studies were conducted by cognitive experiments, for example, by combining two sentences into a single sentence or judging grammaticality of RCs. Those experiments were conducted under a controlled environment. The advantage of such experiments is that a feature effect can be pinpointed and focused, but only very limited amount of samples can be observed and thus might cause incomplete and dispute results.
- In this research we adopted a corpus-based approach. We have the advantages of using a much larger quantity of data from different native language backgrounds and two proficiency levels in order to determine if these groups demonstrate any clear differences when using the RC in Chinese.

TOCFL Learner Corpus

- The TOCFL Learner Corpus consists of essays written by non-native Chinese speaking participants in the Test of Chinese as a Foreign Language (TOCFL) from 2006 to 2012.
- The current TOCFL corpus contains 1.6 million words from learners of 42 different language backgrounds, including 4709 essays on 80 topics written by learners from all proficiency levels.

<http://tocfl.itc.ntnu.edu.tw/> TOCFL Learner Corpus

TOCFL 學習者語料庫檢索系統

說明：「TOCFL 學習者語料庫」所收集的語料是母語非華語的外籍人士參加華語文能力測驗(TOCFL)所寫的作文。寫作時間為2006到2012年。共收集4567篇，約150萬字。其中2837篇(989045字)經過偏誤標記處理。如欲進一步參考語料庫建置內容請參考「[TOCFL作文語料庫的建置與應用](#)」。

中文斷詞及詞性標註採中央研究院CKIP自動斷詞系統。不論自動或人工偏誤標記謬誤難免。開放此系統，僅供研究目的。使用本系統所完成之研究報告或論文，請加註語料出處或致謝。致謝詞範例：本研究感謝國立台灣師範大學TOCFL學習者語料庫團隊提供之語料及檢索系統。

帳號：
密碼：

登入

*如無帳號而欲使用此系統請聯絡：[張老師](#)

2017-5-4

瀏覽人次: 11646

Authentic Materials

- 韩国人在家常吃的菜就是泡菜。(by Korean learner)
- 可是我想还是希望有自己赚的钱。(by Japanese learner)
- 我平常买的菜皆属于不喷农药的「有机蔬菜」。(by English learner)
- 我会先建议你把买回来的那些书利用网络商店转手卖出。(by Spanish)
- 加一点已经炒好的红洋葱会更好吃。(by Indonesian)
- 遇到很多难解决的问题。(by Vietnamese)

Methodology

- In the typological view, we observe six L1s: English, Japanese, Korean, Vietnamese, Indonesian, and Spanish.
- They are grouped into three different typological languages.

Typological groups

	Word order	left or right branching
Japanese, Korean	OV	left-branching (REL-N)
Vietnamese, Indonesian	VO	right-branching (N-REL)
English, Spanish	VO	right-branching (N-REL) and left-branching (adjectival modifiers + head nouns)
Chinese	VO	left-branching (REL-N)

[*xuesheng mai de*] *na ben shu* [学生买的] 那本书 (SVO)

student bought GEN DET CL book

‘the book which the student bought’ (OSV)

• Korean : 학생이 산 책 (SVO)

student-NOM bought book

• Indonesian : buku yang siswa beli (OSV)

book which student buy

• Vietnamese : cuốn sách học sinh mua (OSV)

CL book student buy

• Spanish: El libro que el estudiante compra (OSV)

the book which/that the student bought

Number of Observed Compositions for Six L1s

Compo- sitions	Jap	Eng	Kor	Vie	Indo	Span	Sum
B1 Level	530	344	245	152	163	90	1524
B2 Level	260	122	130	96	112	15	735
Sum	790	466	375	248	275	105	2259

CEFR B1 level: intermediate-high level (ACTFL)

CEFR B2 level: advanced level (ACTFL)

Statistics on RCs

RC tokens	Jap	Kor	Indo	Vie	Eng	Span	Sum
B1 Level	350	299	236	170	239	68	1362
B2 Level	213	146	108	97	107	22	693
Sum	563	445	344	267	346	90	2055

The distribution of subject-RC and object-RC produced by B1 learners

B1 level	Jap	Kor	Indo	Vie	Eng	Span	Average
Sub-RC	24%	21%	22%	24%	23%	31%	23%
Obj-RC	76%	79%	78%	76%	77%	69%	77%

The distribution of subject-RC and object-RC produced by B2 learners

B2 level	Jap	Kor	Indo	Vie	Eng	Span	Average
Sub-RC	47%	59%	25%	35%	60%	5%	45%
Obj-RC	53%	41%	75%	65%	40%	95%	55%

B1 level	Jap	Kor	Indo	Vie	Eng	Span	Average
Sub-RC	24%	21%	22%	24%	23%	31%	23%
Obj-RC	76%	79%	78%	76%	77%	69%	77%

B2 level	Jap	Kor	Indo	Vie	Eng	Span	Average
Sub-RC	47%	59%	25%	35%	60%	5%	45%
Obj-RC	53%	41%	75%	65%	40%	95%	55%

Summary

- Therefore, the investigations of the RCs produced by learners of different native languages and proficiency levels do not support that subject-RCs are easier than object-RCs in Mandarin Chinese.
- On the other hand low-level learners produce more object-RCs consistently. The same result is found in Chinese L1 acquisition research, which indicates the younger the child, the more likely they are to produce an object-RC (Lee, 1992, Cheng, 1995, Chen & Shirai, 2014).
- As a consequence, we hypothesize that object-RC is easier to learn for Chinese because more object-RCs were produced for low level learners regardless their native language types.

RC Position and Type Distribution of B1 Learners

B1 level	Jap	Kor	Indo	Vie	Eng	Span	Average
SS	10%	7.8%	7.7%	4.7%	4.2%	16.2%	44%
SO	38%	39.8%	32.8%	38.2%	34.7%	32.4%	
OS	13.7%	12.6%	14.9%	19.4%	18.92%	14.7%	56%
OO	38.3%	39.8%	45.1%	37.6%	42.2%	36.8%	
Total T	350	294	235	170	239	68	100%

RC Position and Type Distribution of B2 Learners

B1 level	Jap	Kor	Indo	Vie	Eng	Span	Average
SS	16.7%	33.8%	17%	13.7%	16.5%	0%	41%
SO	17.6%	23.3%	36%	23.1%	12.6%	27.3%	
OS	29.9%	24.1%	8%	21.1%	43.7%	4.5%	59%
OO	35.8%	18.8%	39%	42.1%	27.2%	68.2%	
Total T	204	133	100	95	103	22	100%

RC Position and Type Distribution

- The two tables show that the embedded structures of OO and OS are produced more than the non-embedded structures of SS and SO with statistical significance. This violates the theories of language comprehension in cognitive sciences (Bever, 1970, Kuno, 1974, Sheldon 1974). Therefore the results show that easy for comprehension is not equivalent to easy for production in language processing.

The Distribution of Animate Head Nouns Modified by RCs

	B1 Animate Head Nouns		B2 Animate Head Nouns	
	Subject-RC	Object-RC	Subject-RC	Object-RC
English	38% (32)	62% (53)	98% (52)	2% (1)
Japanese	52% (54)	48% (49)	88% (77)	12% (10)
Korean	50% (33)	50% (34)	94% (59)	6% (4)
Indonesian	46% (31)	54% (36)	74% (20)	26% (7)
Vietnamese	77% (24)	23% (7)	85% (23)	15% (4)
Spanish	74% (14)	26% (5)	25% (1)	75% (3)
Average	51% (188)	49% (184)	89% (232)	11% (29)

The Distribution of Animate Head Nouns Modified by RCs

- However for B2 level, totally the Subj-RC (89%) has advantage significantly ($p < 0.0001$) over Obj-RC (11%) in modifying animate head nouns. For the discrepancy between B1 and B2 levels, we will try to explain later in the conclusion.

Conclusion

- The research results show that no matter learners' L1s being left-branching, right-branching, or left-and-right branching, all low proficiency learners produce more Obj-RCs than Subj-RCs. This finding is also true for L1 Chinese acquisition.
- The data leads us to conclude the following hypothesis: At early stage of L2 Chinese learning, word order is the dominant feature for RC processing. Since SVO is conventional word order in Mandarin Chinese and Obj-RC has the same SVO structure. It results that Obj-RCs are easier to learn and thus generated more than subj-RCs.

Conclusion

- This also explains why there is no obvious effect on position feature of causing embedded structure and preference for animate head nouns in Subj-RCs for B1 learners. And after having better language proficiency, the other feature, such as animacy might take over the dominance, though those features are always mingling together to affect the production of RCs.

Limitations

- While a corpus-based study may reflect learner's natural language production, there are also some limits to this research, especially with our hypotheses were established based only on the data of L2 Chinese. We also regret that the results for B2 Spanish learners are just for readers' references since the corpus is not include sufficient materials for producing reliable statistics. Hopefully more corpus data will be produced and sufficient enough for future studies.



Thank
you