Information Status of English Relative Constructions by Japanese Speakers:

A Comparison Between Spoken and Written Language

Megumi Okugiri

Tokyo Healthcare University

m-okugiri@thcu.ac.jp

Although much research has been done in second language (L2) acquisition of relative clauses (RCs) from the viewpoint of syntax (for example, the Noun Phrase Accessibility Hierarchy (NPAH), Comrie & Keenan, 1979) in the past few decades, this topic still provokes controversy. In addition, opinions about the NPAH are divergent with respect to many Asian languages. Teramura (1969, 1993) advocates that there is no such RC in Japanese. Additionally, Comrie (1998, 2002) himself, who proposed the NPAH, later classifies RC-like constructions in Asian languages including Japanese as attributive clauses, based on the argument of Matsumoto (1997), who argues that Japanese attributive clauses are less dependent on the grammatical structure and more dependent on semantic interpretation in context or frame. Comrie (1998, 2002) further proposed that these constructions are structurally different from RCs in European languages, and thus the NPAH may not be the best predictor of the acquisition order of attributive clauses in Asian languages.

In order to clarify and explore this issue, this paper examines the production of English relative constructions of Japanese learners from the discoursal point of view and focuses in particular on the information status of head noun phrases (NPs) modified by RCs.

There are three types of information status, New, Given, and Identifiable[1]. The status refers to the assumptions a speaker makes about whether or not his/her audience is conscious of references that the speaker mentions. The first two, New and Given, were initially proposed by Chafe (1974) in the analysis of NPs in general discourse, and were not limited to relative constructions. Any NPs denote certain information. New information refers to what the addresser assumes not to be in the addressee's

consciousness at all, while Given information is what the addresser assumes to be already present in the current addressee's consciousness. Given information tends to be in pronoun form, e.g., "*I*" in the example shown below (Chafe, 1974, p. 113):

(1) *I* just found *some books* that belong to Peter.

In this example, the lexical NP "*some books*", is a New referent. In addition to these two statuses, Du Bois (1980) proposes the concept of Identifiable referents. An Identifiable referent is presumed not to be in the hearer's present focal consciousness, but the hearer is able to identify the referent as having a strong correlation with a previously introduced referent in terms of frame, prior discourse, or prior knowledge (including world knowledge). Du Bois further argues that identification ordinarily involves singling out the particular referent intended by the speaker. A referent is treated as Identifiable even if it is the first mention as in the following example that the author created:

(2) <u>Jane</u> went to a fancy beauty salon yesterday but didn't get <u>her hair</u> *cut*.

In this example, the information status of *her hair* is Identifiable because it is a part of *Jane*. As in this example, possessive pronouns often allow an object to be marked as identifiable on first mention. The present study will examine these three information statuses of head NPs.

Regarding relative constructions, there are several studies regarding information status in first language (L1). For example, Diessel and Tomasello (2000, 2005) investigated the relative constructions of English-speaking children. They found that the children produced presentational relative constructions (Diessel & Tomasello, 2000, p. 142) that consist of a predicate noun phrase as the head in the main clause and a subject RC as the following example (Diessel & Tomasello, 2000, p. 135):

(3) Is this something that turn around?[2]

A subject RC refers to a RC where its head is the subject of the RC. Diessel and Tomasello propose that these constructions function to draw the children's caretakers' attention to a newly introduced referent as the head NP. Therefore, the head NPs of the presentational relative constructions are generally New referents.

In adult English, Fox and Thompson (1990) investigated the functions of relative constructions in English conversation. They found a tendency for subject RCs to characterise both New and Given head

NPs and for object RCs to identify Given head NPs, and these tendencies are associated with preferences in regard to argument type, i.e., Preferred Argument Structure proposed by Du Bois (1987). Du Bois (2002) examined English native speakers' speech data in spoken American English, analysed all NPs, and proposed a constraint—*the one lexical argument constraint*, which postulates a speaker's preference to avoid more than one lexical core argument in a sentence, as in the following examples (Du Bois 2002: 9):

(4) But *I* enjoyed *the movie*.

(5) *You* called *the police*.

In the examples, *the movie* and *the police* are at the object position, and the arguments take the form of lexical NPs, while the other argument position (i.e., the transitive subject) has only pronouns (i.e., *I* and *you* in the examples). In this way, a speaker maintains the consistency of his/her discourse, introducing only a new single referent at the object position into the discourse.

Studies that discuss the relationship between the information status of NPs and grammatical argument structure include Thompson (1997), who examined lexical NPs and pronouns in each argument position in general English conversation data (including, but not limited to RCs) as a part of her study. Thompson (1997) found that in English an oblique NP is most likely to be New, while a transitive subject NP is least likely to be New.

However, this discoursal preference might vary depending on language. For the information status of NPs in Japanese, Nakayama and Nakayama (1994) investigated the pragmatic properties of NPs in general conversation (including, but not limited to RCs). They found not only a similar tendency for Given transitive subject, but also different frequencies of New object and oblique in comparison to the English results found by Thompson (1997). Nakayama and Nakayama found that Japanese NPs in the object position are most likely to be New. The percentages of New NPs in the studies of Thompson (1997) and Nakayama and Nakayama (1994) are summarised in Table 1. In summary, these studies show that New NPs generally appear in the oblique position in English and in the object position in Japanese.

Table 1: Percentages of New/Given NPs in Japanese and English

	English		Japanese	
	New	Given	New	Given
transitive subject	15 %	85 %	16 %	84 %
intransitive subject	35 %	65 %	22 %	78 %
object	35 %	65 %	57 %	43 %
oblique	65 %	35 %	47 %	53 %

(adapted from Nakayama and Nakayama 1994; 170)

In L2 studies of relative constructions, Okugiri (2012, 2013) demonstrates that semantic and discoursal information affects the mechanisms of the L2 acquisition of the relative constructions in written language. She examined the production of English relative constructions by Japanese learners and native English speakers. The results showed that the learners' preferences of information status were consistent with the discoursal preference of L1 or L2 preference depending on their proficiency. She found that native speakers and advanced learners tend to produce relative constructions when the heads are New as in general English discourse, meanwhile the less advanced learners tend to produce English relative constructions according to the preference of Japanese information status, which is reasonably considered as L1 transfer of discoursal properties, and that the learners produced relative clauses to introduce New referents into the discourse. The present study will examine not only the information status of the head NPs produced by language users at different levels of competence, but also between different modes of language, namely spoken and written. This study will also show that discoursal preference in L1 plays an important role in a learner's spoken language as well as in the written language from the perspective of a usage-based model.

Method

Corpora

The relative constructions of Japanese learners and native English speakers at different levels of competence were extracted from the National Institute of Information and Communications Technology Japanese Learner English Corpus (Izumi, Uchimoto & Isahara, 2005) and the Nagoya Interlanguage

Corpus of English[3] (Sugiura, 2008); the former is a spoken corpus and the latter is a written corpus. Both corpora include data from Japanese learners of English and native speakers of English. The Japanese learners' data for this study were extracted depending on various levels of English proficiency. The learners were grouped into three levels according to their scores on the Test of English for International Communication (TOEIC): a low-intermediate group (scores of 405–600), a high-intermediate group (scores of 605–780), and an advanced group (scores of 785–990).[4] In the spoken corpus, the number of files tagged with the TOEIC score was 123 for the low-intermediate group, 241 for the high-intermediate group, and 194 for the advanced group. For the native group, 20 files were available. Since the author wanted to have more files in order to be able to obtain more conclusive results, the correlation between the TOEIC level and the speaking test level available in the spoken corpus was subsequently analysed. Kendall's rank correlation and the Spearman's rank correlation coefficient were performed. A direct correlation was found between the levels of the TOEIC and the SST: ($W = 0.615$, $p = 0.01$) ($r_s = 0.677$, $p = 0.01$). Thus, the results show that, the higher the TOEIC scores, the higher the speaking test levels. Accordingly, 25 files among the highest level (Level 9) were added to the advanced group. The final numbers of files in the spoken corpus were 123 low-intermediate, 241 high-intermediate, 219 advanced learners, and 20 native speakers of English. The number of files in the written corpus were 37 lower-intermediate, 32 high-intermediate, 25 advanced learners, and 28 native speakers of English.

The relative constructions were initially gathered by extracting relative pronouns (i.e., *that, which, who, whom, and whose)*[5]. Relative constructions without relative pronouns (e.g., the woman I know) were extracted by hand.

Definitions of the Information Status of the Head

The definitions for the information status of heads are those used by Fox and Thompson (1990), which was originally based on the information status proposed by Chafe (1974) and Du Bois (1980): *New, Given* and *Identifiable*. As explained earlier, a New referent is newly introduced into the discourse and presumed not to be in the addressers focal consciousness or understanding. A Given referent is presumed to be in the addressers focal consciousness and is distinguished as a referent that is anaphorically linked to

its previous mention. An Identifiable referent is presumed not to be the addressers present focal consciousness, but the addresser is able to identify the referent as having a strong correlation with a previously introduced referent in terms of frame, prior discourse or prior knowledge (including world knowledge).

Head Types

Head types were determined depending on the position of the head NPs. There are seven types: intransitive subject (S), transitive subject (A), object (O), oblique (B), main clause (M) where the head is the whole main clause, noun phrase (NP) and predicate nominal (PN). Examples are shown below:

Intransitive subject (S): *The man* [who has the dog] is Mary's brother.

Transitive subject (A): *The man* [who has the dog] has a lot of money.

Object (O): He saw *the woman* [who has the dog].

Oblique (B): He lives with *the woman* [who works for a big company].

Main clause (M): *He was diabetic*, [which I didn't even know that cats could get].

Noun phrase (NP): *The man* [who has the dog].

Predicate nominal (PN): There is *a man* [who has a dog].

The heads are in italic and the RCs are in brackets.

Results and Discusssion

In total, 1,988 relative constructions were found in the corpora. Table 2 shows the frequency of information status of the head NPs in each group and mode. The bold numbers indicate the type of the most frequent information status. The results show that both learners and native speakers produce New heads the most frequently in both spoken and written modes, and a chi-square analysis yielded no significant difference between the information status and proficiency levels (χ^2 (6, N = 1,998) = 6.114, p = .4106 *n.s.*).

Table 2

Frequency of the Information Status of the Head Noun Phrases

Level	Mode	New	Given	Identifiable	Total
Low-Intermediate	spoken	**28 (58.33%)**	10 (20.83%)	10 (20.83%)	48 (100%)
	written	**32 (50.00%)**	10 (15.63%)	22 (34.38%)	64 (100%)
High-Intermediate	spoken	**151 (54.12%)**	47 (16.85%)	81 (29.03%)	279 (100%)
	written	**58 (55.24%)**	19 (18.10%)	28 (26.67%)	105 (100%)
Advanced	spoken	**348 (54.63%)**	99 (15.54%)	190 (29.83%)	637 (100%)
	written	**65 (50.00%)**	26 (20.00%)	39 (30.00%)	130 (100%)
Native	spoken	**341 (64.71%)**	62 (11.76%)	124 (23.53%)	527 (100%)
	written	**96 (46.15%)**	48 (23.08%)	64 (30.77%)	208 (100%)
Total		**1,119**	321	558	1,998

Table 3 shows the frequency of the head types in each group and mode. The bold numbers indicate the most frequent types, and the numbers shaded in gray indicate relatively frequent head types. The token displayed in the rightmost row is the frequency per file (namely participant). The token increases as the learners become more advanced.

Table 3

Frequency of the Head Types

Level	Mode	S	A	O	B	M	NP	PN	Total	Token
Low-Intermediate	spoken	5 10.42%	0 0.00%	**23** **47.92%**	10 20.83%	0 0.00%	1 2.08%	9 18.75%	48 100%	0.39
	written	6 9.38%	7 10.94%	16 25.00%	**17** **26.56%**	2 3.13%	2 3.13%	14 21.88%	64 100%	1.72
High-Intermediate	spoken	33 11.83%	11 3.94%	**93** **33.33%**	59 21.15%	0 0.00%	23 8.24%	60 21.51%	279 100%	1.16
	written	**29** **27.62%**	8 7.62%	26 24.76%	20 19.05%	0 0.00%	5 4.76%	17 16.19%	105 100%	3.28
Advanced	spoken	78 12.24%	33 5.18%	**172** **27.00%**	167 26.22%	10 1.57%	23 3.61%	154 24.18%	637 100%	2.91
	written	18 13.85%	6 4.62%	23 17.69%	38 29.23%	3 2.31%	2 1.54%	**40** **30.77%**	130 100%	5.2
Native	spoken	42 7.97%	7 1.33%	131 24.86%	122 23.15%	48 9.11%	39 7.40%	**139** **26.38%**	527 100%	26.35
	written	41 19.71%	10 4.81%	36 17.31%	**70** **33.65%**	3 1.44%	2 0.96%	45 21.63%	208 100%	7.43
Total		252	82	519	502	66	25	552	1,998	

S=intransitive subject, A=transitive subject, O=object, B=oblique, M=main clause, NP=noun phrase, PN=predicate noun phrase

The frequent heads were in the low-intermediate group, object head for the spoken mode, and object, oblique and PN heads for the written mode; in the high-intermediate group, object head for the spoken mode, and intransitive subject and object heads for the written mode; in the advanced group, object in the spoken mode and oblique and PN heads for both modes; and in the native group, PN, object and oblique heads for the spoken mode, and oblique head for the written mode. The overall results supported the findings of Thompson (1997) and Nakayama and Nakayama (1994).

Chi-square analyses were performed to compare the results in each mode and group. And the results yielded a significant difference between the spoken and written data of the high-intermediate group (χ^2 (5, N = 384) = 18.181, p = .0027*) and of the native group (χ^2 (6, N = 735) = 62.019, p < .0001*). The results of the low-intermediate (χ^2 (5, N = 112) = 11.532, p = .0733 $n.s.$) and of the advanced (χ^2 (6, N = 767) =7.961, p = .2410 $n.s.$) groups were not significant. Therefore, the statistical results indicate that the results in the spoken mode were significantly different from those of the written mode in the high-intermediate group and the native groups, while the production of relative constructions in the spoken and written modes were not much different in the low-intermediate and advanced groups.

Nevertheless, it is notable that the high-intermediate group produced the intransitive subject head the most frequently in the written mode, which is not the case in the native group. Thus, the statistical significance of the high-intermediate group does not imply that their production was similar to the natives; rather, their production seemed to be idiosyncratic among the learner groups. As for the results of the high-intermediate group, I reexamined the data in terms of animacy status and found the frequent intransitive subject head is due to the learners' tendency to produce RCs when they refer to human referents. This phenomenon has been discussed in Okugiri (2012) and will not be explained here further due to space constraints.

Regarding the results of the spoken mode in the native group, the most frequent head type in the native group was PN. This result is in line with the finding by Diessel and Tomasello (2000) in the production of children's English relative constructions, i.e., presentational relative constructions. The pragmatic function of PN heads to introduce New referents into the discourse is probably more crucial in the spoken mode among native speakers.

The learners also frequently produced PN heads, but in the written mode (21.88% for Low-Intermediate, 16.19% for High-Intermediate, and 30.77% for Advanced), which may be a result of their intention to introduce New referents into the discourse, something which they were not able to manage easily with less planning time in the spoken mode. As such, they may depend more on the Japanese preference, which is more easily accessible when they are affected by real-time processing pressure in a spontaneous-speaking mode.

The results reveal that the preference of the L1 Japanese had a strong influence on the learners' production of English relative constructions and that the influence was even stronger among the less advanced learners and in the spoken mode than in the written mode. The different results between the two modes seem to be due to less planning time in the spontaneous-speech mode (Ochs, 1979; Skehan, 1998). However, further research is required to determine whether or not planning time actually affects production in the two different modes.

Conclusion

The results of this study showed different patterns of relative constructions between learners and native speakers depending on the learners' proficiency and on the mode of language, the influence of the preference of L1 information status on the acquisition of L2 English relative constructions, and its stronger influence in the spoken mode of the less advanced learners. This study suggests that discoursal properties play a major role in acquisition, which reflects learners' cognitive processing of a second language.

Notes

[1] The terms are capitalised in order to avoid confusion with other uses in expository English and to indicate them as information status terms.
[2] The expected sentence is, "Is this something that turns around?"
[3] This study used an older version of the Nagoya Interlanguage Corpus of English because it was the latest version available when the data was extracted for this study. The most current version includes a larger number of files and is now available at http://sgr.gsid.nagoya-u.ac.jp/wordpress/?page_id=17.
[4] These levels are based on the Can-Do Levels Table provided by the Educational Testing Service (http://www.uk.etseurope.org/ home-corpo-uk/toeic-can-do-table/).
[5] Relative sentences with the relative pronouns *where* or *what* are not included in this data because this paper

focuses on RCs that require a head NP. *Where*, which is an adverbial relative pronoun, frequently does not require an overt head NP, and *what*, which is a nominal relative pronoun, never allows an overt head NP.

References

Chafe, W. (1974). Language and consciousness. *Language, 50* (1), 111-133.

Comrie, B. (1998). Attributive clauses in Asian languages: Towards an areal typology. In W. Boeder, C. Schroeder, K. H. Wagner & W. Wildgen (Eds.), *Sprache in Raum und Zeit: In memoriam Johannes Bechert* (2, pp. 51-60). Tübingen: Gunter Narr.

Comrie, B. (2002). Typology and language acquisition: The case of relative clauses. In A. G. Ramat (Ed.), *Typology and Second Language Acquisition* (pp. 19-37). Berlin: Mouton de Gruyter.

Comrie, B., & Keenan, E. L. (1979). Noun phrase accessibility revisited. *Language, 55* (3), 649-664.

Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics, 11*, 131-151.

Du Bois, J. (1980). Beyond definiteness: The trace of identity in discourse. In W. Chafe (Ed.), *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production* (pp. 203-274). Norwood, NJ: Ablex.

Du Bois, J. (1987). The discourse basis of ergativity. *Language, 63* (4), 805-855.

Du Bois, J. (2002). Discourse and grammar. In Michael Tomasello (Ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure* (Vol. 2, pp. 47-88). Mahwah, NJ: Lawrence Erlbaum.

Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in

English conversation. *Language, 66* (2), 297-316.

Izumi, E., Uchimoto, K., & Isahara, H. (2005). *The National Institute of Information and Communications Technology Japanese Learner English Corpus*. Tokyo: ALC Publishing.

Matsumoto, Y. (1997). *Noun-Modifying Constructions in Japanese: A Frame-Semantic Approach.* Amsterdam: John Benjamins.

Nakayama, T., & Nakayama, K. (1994). Discourse perspective for core-oblique distinction in Japanese. *Santa Barbara Papers in Linguistics, 5*, 158-177.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givon (Ed.), *Discourse and Syntax* (pp. 51-80). New York: Academic Press.

Okugiri, M. (2012). The animacy status and the acquisition of English relative clauses in written language by Japanese learners. In Tokyo Daigaku Paul Rossiter Kyōju Taikan Kinen Syuppan Iinkai (Ed.), *West to East, East to West: Studies in the Field of English Education Dedicated to Professor Rossiter on His Retirement* (pp. 135-156). Tokyo: Seibidō.

Okugiri, M. (2013). The use of English relative constructions by Japanese learners in written language. *The International Journal of Literacies*, *19*(13), 13-24.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Sugiura, M. (2008). Eigo gakusyūsya no korokēsyon chishiki ni kansuru kisoteki kenkyū [Nagoya Interlanguage Corpus of English (NICE)]. Retrieved October 5, 2008, from http://sugiura5. gsid.nagoya-u.ac.jp/~sakaue/nice/index.html

Teramura, H. (1969). The syntax of noun modification in Japanese. *The Journal-Newsletter of the Association of Teachers of Japanese, 6* (1), 64-74.

Teramura, H. (1993). *Teramura Hideo Ronbunshū I: Nihongo Bunpō Hen* [Collection of Papers by Hideo Teramura, Vol. 1: Japanese Syntax] (reprinted in 1999). Tokyo: Kuroshio.

Thompson, S. A. (1997). Discourse motivations for the core-oblique distinction as a language universal. In A. Kamio (Ed.), *Directions in Functional Linguistics* (pp. 59-82). Amsterdam: John Benjamins.

＜要旨＞

日本語母語話者による英語関係節構文の習得：
話し言葉と書き言葉における情報の新旧の比較

奥切　恵

東京医療保健大学

m-okugiri@thcu.ac.jp

　本稿は、日本語母語話者の英語関係節構文の習得を談話的観点から考察し、談話特徴が関係節構文の習得において大きく影響していることを明らかにする。過去の関係節構文の習得研究は、その統語要素のみに特化した研究がほとんどで、談話の観点から研究したものは少ない。本研究では、学習者による英語関係節構文の談話特徴を習熟度別に分析し、話し言葉と書き言葉、また英語母語話者と学習者の産出の比較をした。データは、学習者の話し言葉コーパスからは123名の初中級学習者、241名の中級学習者、219名の上級学習者、20名の英語母語話者、書き言葉コーパスからは37名の初中級学習者、32名の中級学習者、25名の上級学習者、28名の英語母語話者、以上合計725名のファイルから関係節構文を抽出した。抽出された関係節構文の数は、合計1998文であった。その関係節構文の先行詞の情報の新旧を新情報、同定可能情報、旧情報の3つに分類し、分析をした。

　分析の結果では、学習者と英語母語話者の間に統計的に有意な差がみられた。英語母語話者は一般的な英語での談話内にみられるように、関係節を使用する際にも斜格名詞句に新情報をおき、関係節の先行詞とする傾向がみられた。一方学習者においては、特に習熟度の低い学習者は一般的な日本語の談話内でみられるように、目的格名詞に新情報をおき、英語関係節の先行詞とする傾向が強いことがわかった。本研究により、第一言語習得だけでなく第二言語習得においても、情報の新旧といった談話特徴が、関係節構文の産出に大きく影響していることが明らかとなった