

Research article

Frequency and Complexity of Relative Clauses in Vietnamese Learners' Essays

Vo Thanh Nga*, Dang Hoai Phuong

International University, VNU-HCMC

ORCIDVo Thanh Nga: <https://orcid.org/0000-0001-5130-6649>**Abstract.**

Relative clauses are considered as complex structures and employed in academic writing to a high extent. The frequency and complexity of these structures is expected to make non-native learners' writing more academic. Therefore, the paper aims to investigate how Vietnamese learners of English use relative clauses in their essays. Particularly, 100 argumentative essays written by intermediate students are examined to explore the frequency and complexity of these structures. The findings show that the presence of relative clauses is found in nearly all essays. However, the investigated students tend to use many more restrictive relative clauses than non-restrictive ones. Regarding syntactic complexity, they have a preference order of SU>DO>OBL>GEN>IO=OCOMP. It is suggested that non-native learners should acquire the use of all types of relative clauses for them to be appropriately employed in their academic writing.

Corresponding Author: Vo Thanh Nga; email: vtnga@hcmiu.edu.vn

Dates

?????

Publishing services provided by Knowledge E

Keywords: relative clauses, academic essays, non-native users, the NPAH

© Vo Thanh Nga, Dang Hoai Phuong. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Selection and Peer-review under the responsibility of the iNETAL Conference Committee.

1. Introduction

Many researchers believe that syntactic complexity is a reliable indicator for linguistic proficiency. According to Hunt [1], the length of a T-unit including a main clause and embedded clauses, has an interrelationship with the quality of writing. In this regard, relative clause (RC), which is defined as a "subordinate clause that modifies a noun or noun phrase in an associated main clause" [2], has attracted interest of many researchers as well as writing teachers. In fact, promoting the proper use of RCs by language learners has been considered a way to improve their quality of writing.

There are various classifications of RCs. Based on the necessity of the modification, RCs can be divided into restrictive and non-restrictive ones [3]. A restrictive RC provides information that is essential for defining the head noun while a non-restrictive RC only gives extra information to it and is often separated from the head noun with a comma [4]. The difference between these two types of RCs can be shown in the following examples:

OPEN ACCESS

(a) Restrictive clause: *The student who sits next to the window does not come to class today.*

(b) Non-restrictive clause: *John, who sits next to the window, does not come to class today.*

Based on the syntactic function of the head noun within the RC, Keenan and Comrie [5] categorized RCs into 6 types that fall within a universal hierarchy. Investigating RCs in more than 50 different languages, the two language researchers outlined the Noun Phrase Accessibility Hierarchy (NPAH), suggesting an order to the possibility of RC formation in almost all languages. The hierarchy can be illustrated as follows:

Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of comparison

(SU > DO > IO > OBL > GEN > OCOMP)

According to the NPAH, the most accessible position for relativization is the subject and the least accessible one is the object of comparison. It implies that if a certain position in the hierarchy can be relativized, RCs can also be formed on all the higher positions to its left. Despite the limitation that the hierarchy does not include an RC at the adverb position [6], the NPAH is still central in most research into the acquisition of RC and will be used as the framework for this study.

Based on the syntactic role of the head noun in the matrix clause and the syntactic role of the gap or the relativized constituent within the relative clause, RCs can also be categorized into four types including SS, SO, OS and OO.

SS	The head noun is the subject of the matrix clause, and the gap is the subject of the RC. Example: <i>The man who is talking with our teacher is my uncle.</i>
SO	The head noun is the subject of the matrix clause, and the gap is the object of the RC. Example: <i>The man whom you met yesterday is my uncle.</i>
OS	The head noun is the object of the matrix clause, and the gap is the subject of the RC. Example: <i>I know the man who is talking with our teacher.</i>
OO	The head noun is the object of the matrix clause, and the gap is the object of the RC. Example: <i>I know the man whom you met yesterday.</i>

Figure 1: The four types of relative clauses.

Despite being a universal phenomenon, these syntactically complex structures vary from language to language, which makes RCs one of the most challenging structures to language learners. Therefore, it is necessary to investigate the acquisition of RCs by L2 learners. There are two clear trends of research on RCs. The first trend focuses on

testing different hypotheses on the difficulty order of RCs. The four main hypotheses that are commonly tested are the NPAH [5], and the other three hypotheses by Kuno [7], Keenan [8] & Sheldon [9] regarding the four aforementioned types of RCs - SO, SS, OS, OO. Using different types of tests and learners' compositions, researchers counted the frequency of RCs used and the total number of errors to confirm or reject the hypotheses proposed by the previous scholars [10, 11, 12, 13]. The second line of research is concerned with the types and sources of errors in the use of RCs by L2 learners [11, 14, 15, 16, 17]. The most common types of errors are wrong use of relative pronouns, inappropriate deletion of relative pronouns, absence of an antecedent, use of resumptive pronouns, disagreement between subject and verb, unnecessary use of RCs and mismatch between L1 and L2 head noun direction. It is also found that L1 transfer and intralingual factors are the main causes of errors.

However, very few studies have been done on the use of RCs by Vietnamese learners, especially in writing. Therefore, this study, in the hope of contributing to the field and suggesting possible pedagogical implications, aims at finding out how Vietnamese learners use RCs in their writing and the types of errors that are frequently made. With these purposes, the study attempts to answer the following questions:

1. To what extent do Vietnamese learners of English employ RCs in their argumentative essays?
2. What types of RCs are frequently used in their argumentative essays based on the NPAH [5]?
3. What types of errors regarding RCs may the students make in their essays?

2. Method

2.1. Data

As described in Table 1, the data were collected from 100 argumentative essays written by Vietnamese learners of English whose L2 proficiency level is intermediate. The average word number for each essay is 383 words, ranging from 199 to 666 words. As a result, the size of the data is 38,303 words.

TABLE 1: Description of the data.

Variable	N	Mean	Sum	Minimum	Maximum
Word number	100	383.03	38303.00	199.00	666.00

2.2. Data analysis

The study was conducted in both the qualitative and quantitative manner. First, the RCs employed in the students' argumentative essays were identified and classified into the categories of restrictive RCs and non-restrictive RCs as well as those of SU, DO, IO, OBL, GEN and OCOMP based on the NPAH [5]. The erroneous structures were also recorded. Second, the frequencies of RC instances in each category were measured to answer the first two research questions. Finally, the erroneous RCs were examined carefully to figure out the error types students may frequently encounter in the use of RCs.

3. Findings and Discussion

3.1. RC Frequency

The frequency of RCs employed in the argumentative essays written by Vietnamese L2 learners is described in Table 2.

TABLE 2: RC frequency.

Variable	N	Mean	Sum	Minimum	Maximum
RC number	100	2.930	293.000	0.000	12.000

In general, 293 RCs were identified in 100 argumentative essays; particularly, almost 3 instances were averagely employed in each of them, and the frequency ranges from 0 to 12 RCs. This means the current research found 7.6 RCs per 1000 words. The findings in other studies are varied. Shahriari [18] found 12 RCs per 1000 words in the corpus of argumentative essays by Iranian EFL learners and 7 RCs per 1000 words in the corpus by native writers. Biber et al. [4] found 9.9 RCs per 1000 words in the corpus of academic prose including book extracts and scientific papers. Cho & Lee [19] found from 5 to 8.8 RCs per 1000 words in the four investigated journals. Deveci & Nunn [20] found 17 RCs per 1000 words in the corpus of research articles in five journals. The findings reflect that RCs are employed in academic writing and its use contributes to enhancing the academic standards of the written texts. The difference in RC frequency may be explained by the variation in the investigated RC users and text types.

3.2. RC Complexity

A closer look at the proportion between restrictive and non-restrictive RCs reveals that the number of the former ones significantly surpasses that of the latter ones. Specifically, as described in Figure 2, 261 restrictive instances which take up 89% of the total were recorded in comparison with 32 non-restrictive RCs (11%).

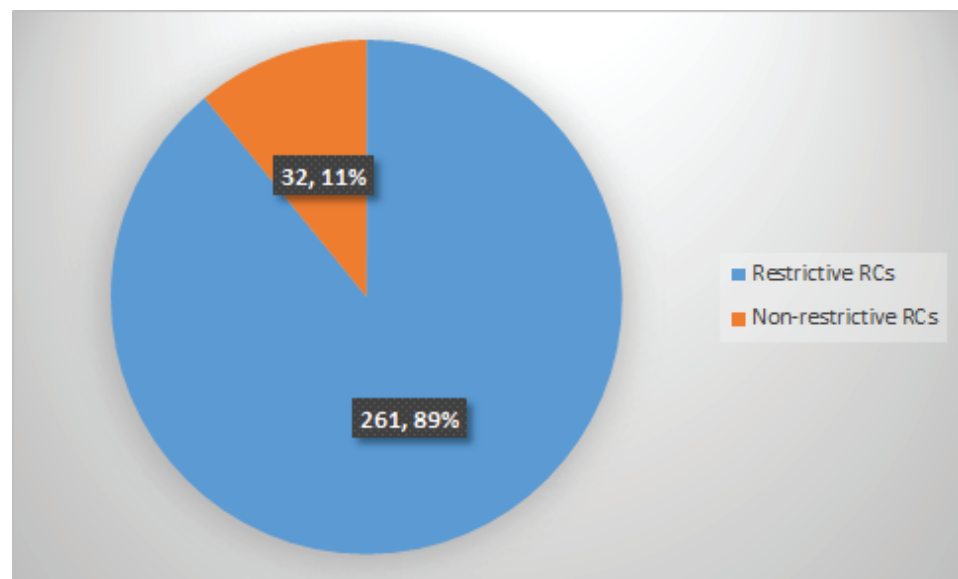


Figure 2: Restrictive and non-restrictive RCs.

The asymmetry between restrictive and non-restrictive RCs is also confirmed in previous studies. Biber et. al. [4] found the proportion of 77% restrictive RCs and 23% non-restrictive ones. Also, Shahriari [18] found the proportion of 88.79% and 11.20% in L2 learners' essays as well as that of 91.97% and 8.02% in L1 writers' essays. Nevertheless, it is interesting that the findings by Cho & Lee [19] confirm a wide gap between these two RC types in two investigated journals but shows a narrow gap in the other two journals of chemistry and electrical engineering which require more explanations for unfamiliar concepts. Therefore, it can be concluded that restrictive RCs are much more common than non-restrictive ones and their asymmetric distribution depends on the written texts.

Furthermore, these RCs are also classified into the six types according to the NPAH [5]. Figure 3 below illustrates the distribution of the identified RCs in the categories of SU, DO, IO, OBL, GEN and OCOMP.

It is observed that the SU structures are the majority among the six types of RCs; they take up 80% of all the identified RCs. The second highest ones are DO structures whose frequency is 53 (18%). The third and fourth ranks are for OBL and GEN structures;

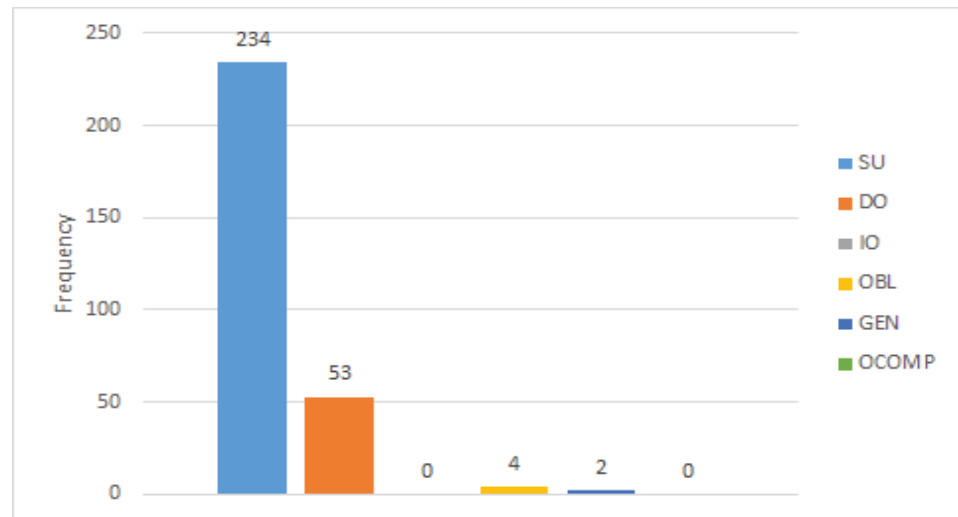


Figure 3: RCs based on the NPAH [5].

the significantly low figures are 4 and 2 respectively. None of the IO and OCOMP structures is present in the data. Therefore, the RC hierarchy reflected in the data is SU>DO>OBL>GEN>IO=OCOMP.

The finding that SU RCs are the most frequent and easiest structures for EFL learners is consistent with those in previous research [21, 14, 22, 17]. Besides, it is confirmed that GEN and OBL structures cause more difficulties for L2 students [21, 14, 13]. The hierarchy found in the current study confirms the NPAH except for the IO structures, none of which was recorded in the data. The missing of IO structures in learners’ essays may be attributed to the small scope of this study rather than an implication of difficulty.

3.3. RC Errors

More details regarding how the L2 learners used RCs in their argumentative essays are reflected in Table 3 below. It is observed that they may form incorrect RCs in an attempt to employ these complex structures in their writings.

TABLE 3: Correct and incorrect RCs based on the NPAH [5].

Types	#RCs	Correct RCs	Incorrect RCs
SU	234	190 (81%)	44 (19%)
DO	53	45 (85%)	8 (15%)
IO	0	0 (0%)	0 (0%)
OBL	4	3 (75%)	1 (25%)
GEN	2	1 (50%)	1 (50%)
OCOMP	0	0 (0%)	0 (0%)
Total	293	239 (81%)	54 (19%)

In general, there are 54 incorrect RC instances which take up 19% of all the identified RCs. This shows that RCs still cause certain problems for students who employ them to enhance the academic writing standards. In particular, 19% of the SU structures and 15% of the DO ones are erroneous. The significant finding is the high percentages of incorrect instances in the OBL and GEN categories (25% and 50% respectively). This may confirm their standings on the NPAH, i.e., they are less common in L2 learners' essays, and these students find them more structurally complicated.

An examination of the errors the Vietnamese students encounter shows a detailed picture of their RC use.

TABLE 4: RC errors.

Error Type	Frequency	Percentage
Subject-verb disagreement	20	32
Use of the resumptive pronoun	1	2
Wrong use of the relative pronoun	4	7
Inappropriate ellipsis of the relative pronoun	16	26
Absence of the antecedent	3	5
Unnecessary use of RCs	8	13
Others	9	15
Total	61	100

Overall, 61 RC errors were identified and classified into 7 categories, as described in Table 4. It is noted that there are 54 incorrect RCs but 61 errors because some RCs have multiple errors. For example, the RC *contact with whom not having vaccination* may contain three error types including the absence of the antecedent, the wrong use of the relative pronoun and the absence of a finite verb. The correct version should be *contact with those who do not have vaccination*.

Specifically, the two most frequent mistakes by the Vietnamese students involve subject-verb disagreement and relative pronoun absence. 32% and 26% of the total errors fall into these error categories. The examples for these errors are shown below; the subject-verb mismatch is illustrated in (1) and (2), and the relative noun deletion is exemplified in (3) and (4). It is noted that the asterisk signals incorrect RCs and their revised versions are also suggested.

(1) Incorrect RC : **many new variants which is more dangerous*

Revised RC : *many new variants which are more dangerous*

(2) Incorrect RC : **the company which produce the vaccine*

Revised RC : *the company which produces the vaccine*

(3) Incorrect RC : **there are some ideas said that after vaccination*

Revised RC : *there are some ideas which said that after vaccination*

(4) Incorrect RC : **there are many people say that*

Revised RC : *there are many people who say that*

In addition, the error of unnecessary use of RCs which takes up 8% also needs examining. The following structure exemplifies the unnecessarily complicated use of RCs.

(5) *There are many people who contract the disease outside the community and spread it to family members that include the elderly and children who have weak health and low resistance.*

Example (5) contains three RCs which are correctly structured, but they all make the whole sentence complicated. Therefore, their use seems to be redundant, and the writer may need to revise this stringy sentence.

The other two examples for the unnecessary use of RCs are as follows:

(6) *my neutral opinion which is both agree and disagree with it*

(7) *the secret nobody can know*

In (6) and (7), the antecedents can express the meanings of the whole phrases without the RCs. Therefore, these RCs seem to be unnecessary.

Furthermore, the three errors in which relative pronouns are incorrectly used, antecedents are absent and resumptive pronouns are unnecessarily employed take up 7%, 5% and 2% respectively. These are minor in comparison with the previously mentioned three error types.

Examples (8) and (9) illustrate the wrong use of relative pronouns.

(8) **the pandemic of Covid 19, that breaks out in the two-year period*

(9) **contact with whom not having vaccination*

In (8), the relative pronoun *which* should be used instead of *that* because the non-restrictive RCs follow a comma. Besides, as mentioned earlier, in (9), the antecedent is absent, and then *who* should be used instead of *whom*.

The error of antecedent absence is shown in (10).

(10) **the number of who died from this pandemic*

The antecedent *people* may be added before the relative pronoun to make the clause correct.

The unnecessary use of a resumptive pronoun is in (11).

(11) **a right and obligation of every citizen that they have no right*

The best revised version for this clause is *a right and obligation which every citizen does not have*. Therefore, it may be seen that the pronoun *they* as well as the noun *right* should be deleted.

It is also interesting to mention the last category titled *Others*. These incorrect RCs which take up 15% may not contain RC-related errors. Instead, they involve grammatical errors in which main finite verbs are not present. Some examples are (12), (13) and (14).

(12) Incorrect RC : **vaccines that against covid*

Revised RC : *vaccines that are against covid*

(13) Incorrect RC : **people who not having vaccination*

Revised RC : *people who do not have vaccination*

(14) Incorrect RC : **many benefits that mentioned above*

Revised RC : *many benefits that are mentioned above*

It is seen that the clauses are grammatical when finite verbs are added.

4. Conclusions

This study investigated the use of RCs in argumentative essays by Vietnamese intermediate learners. In general, RCs were found to be employed by these L2 learners to enhance their essays' academic quality. Furthermore, the results corroborated the NPAH by Keenan & Comrie [5] in that subject and direct object were the most accessible positions to be relativized while RCs at the lower positions in the hierarchy were less frequently used. It was also found that restrictive RCs were more common than non-restrictive clauses, which is in line with the previous studies [4, 19, 18]. The qualitative analysis of RCs revealed a range of errors made by the Vietnamese intermediate students, among which head-verb disagreement and omission of relative pronouns were prominent. Despite the small size of data, the findings can shed light on the teaching of RCs. It is suggested that teachers familiarize students with various types of RCs in the NPAH by providing diverse examples and exercises. Also, the common types of errors with RCs should be indicated and emphasized in lessons to help students avoid them.

References

- [1] Hunt, K. W. Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development*. 1970;35(1):iii-67. doi:10.2307/1165818

- [2] Diessel, H., & Tomasello, M. The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*. 2001;11:131-151. doi:10.1515/cogl.2001.006
- [3] Lock, G. *Functional English Grammar: An introduction for second language teachers*. Cambridge: Cambridge university press; 1996.
- [4] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. *Grammar of spoken and written English*. Essex: Pearson Education Limited; 1999.
- [5] Keenan, E., & Comrie, B. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*. 1977;8(1):63-99.
- [6] Phoocharoensil, S. Errors on the relative marker WHERE: Evidence from an EFL Learner Corpus. *The Southeast Asian Journal of English Language Studies*. 2014;20(1):1–20. <http://dx.doi.org/10.17576/3L-2014-2001-01>
- [7] Kuno, S. The position of relative clauses and conjunctions. *Linguistic Inquiry*. 1974;5:117-136.
- [8] Keenan, E.L. Variation in universal grammar. In R. W. Fasold and R. W. Shuy (eds.), *Analysing variation in language*. Washington DC: Georgetown University Press; 1975.
- [9] Sheldon, A. The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*. 1974;13(3):272-281.
- [10] Wong, J. Learnability of relative clauses: A Hong Kong Case. *Working Papers of the Department of English*. 1991;3(1):108-117.
- [11] Chang, Y. Second language relative clause acquisition: An examination of cross-linguistic influences. Paper presented at the Annual Meeting of the American Association for Applied Linguistics; 1-4 May 2004; Portland.
- [12] Abdolmanafi, S., & Rahmani, Z. An Investigation of the Learnability of Relative Clauses by EFL Learners. *World Journal of English Language*. 2012;2(3):29-37.
- [13] Hassan, I. H. Investigating the acquisition order of English restrictive relative clauses: Evidence from Egyptian EFL Learners. *Journal of the Faculty of Education at Alexandria University*. 2018;28(1):313-343.
- [14] Phoocharoensil, S., & Simargool, N. English Relativization and Learners' Problems. *Journal of Pan-Pacific Association of Applied Linguistics*. 2010;14(1):109-129.
- [15] Bao, X. An investigation of sources of errors in the use of relative clauses by English majors. *English Language and Literature Studies*. 2015;5(3):137-145.
- [16] Novianto, W. & Suhandano, S. Production of relative clauses by English learners. *Proceedings of English Language and Literature International Conference*, 2019;3:50-56.

- [17] Zhu, F. Restrictive relative clauses in EFL writing. *OnCUE Journal Special Issue*. 2020;2:84-92.
- [18] Shahriari, H., Shadloo, F. & Ansarifar, A. An examination of relative clauses in argumentative essays written by EFL learners. *Journal of Language and Education*. 2018;4(4):77-87. <https://doi.org/10.17323/2411-7390-2018-4-4-77-87>
- [19] Cho, D. W. & Lee, K. English relative clauses in science and engineering journal papers: A comparative corpus-based study for pedagogical purposes. *Ampersand*. 2016;3:61-70.
- [20] Deveci, T. & Nunn, R. The use of relative clauses in humanities and social sciences research articles: A Case Study. *Linguistics and Literature Studies*. 2018;6(1):17-26.
- [21] Kubota, M. Accuracy Order and Frequency Order of Relative Clauses as Used by Japanese Senior High School Students of EFL. *Institute for Research in Language Teaching Bulletin*. 1993;7:27-53.
- [22] Alotaibi, A. M. Examining the learnability of English relative clauses: Evidence from Kuwaiti EFL learners. *English Language Teaching*. 2016;9(2):57-65.