

## ON THE OMISSION OF RELATIVE PRONOUNS AMONG CHINESE SPEAKERS OF ENGLISH IN ACADEMIC WRITING

Hsuan Yu Tai

### ABSTRACT

The use of relative pronouns is an intriguing topic for learners and teachers. The past two decades have witnessed exponential growth in research on the use of relative pronouns. However, little attention has been given to relativizer omission among nonnative speakers in writing. This research examined the distribution of zero relativizers and the variation patterns of relative pronoun omission in academic writing among Chinese speakers of English. The data were collected from 396 written essays in two English-as-a-medium-of-instruction courses. Adopting a variationist approach, the written essays were analyzed by the Variable Rule (VARBRUL) analysis, a multivariate logistic regression. The findings revealed that through particular essay sentence structures, syntactic functions, type of antecedents, and length of relative clauses, one could predict the writer's choice of zero relative pronouns. The reasons for variation can be explained by the effects of language processing, predictability or frequency. Some pedagogical implications are proposed.

**Key Words:** language variation, zero relativizer, relative pronoun omission, Chinese learners of English

### INTRODUCTION

Relative pronouns are a significant topic for academics in linguistics (Bayley, 1999; Guy & Bayley, 1995; Grafmiller, Szmrecsanyi & Hinrichs, 2018; Levey, 2006; Levey & Hill, 2013). English speakers need to select from the relative pronouns of *who*, *which*, *that* and zero forms (Bayley, 1999; Guy & Bayley, 1995; Levey & Hill, 2013). Given that relative pronouns can be used interchangeably under certain circumstances, the four different relative pronouns are, from the variationist's perspective, variants. To uncover the reasons for relative pronoun variation, some

scholars have explored the changes of relative pronouns from the diachronic point of view (Ball, 1994, 1996). Conversely, other researchers have attempted to understand relative pronouns through investigating their rules of variation. In other words, given that the choices of relative pronouns can be attributed to various linguistic factors, scholars (Bayley, 1999; Guy & Bayley, 1995; Levey & Hill, 2013) attempted to include contributing linguistic and extra-linguistic factors in the analysis and employed Variable Rule (VARBRUL) analysis (Sankoff, 1988b), a multivariate logistic regression, to create a parsimonious model which could predict the use of relative pronouns.

In terms of the research on the rules of variation in relative pronoun omission, previous findings have been quite fruitful. Researchers have explored the systematic patterns of varieties in different speech communities. These communities include speakers in America (Guy & Bayley, 1995), the U.K. (Levey, 2006; Tagliamonte, Smith & Lawrence, 2005), and others in the Outer Circle (Rickford, 2011). However, scant research has been collected on nonnative speakers of English (Flanigan & Inal, 1996).

From the research studies on the variation in relative pronoun omission, researchers have shown that certain linguistic or psycholinguistic factors could affect relative pronoun omission simultaneously and to different degrees. These factors include syntactic functions (Guy & Bayley, 1995), sentence structures (Levey & Hill, 2013), animacy of antecedents (Jeager & Wasow, 2006), and length of relative clauses (Grafmiller, Szmrecsanyi & Hinrichs, 2018). Speech corpora provided the basis for the majority of the research (Bayley, 1999; Guy & Bayley, 1995; Levey, 2006; Levey & Hill, 2013; Rickford, 2011; Tagliamonte, Smith & Lawrence, 2005; Tottie & Rey, 1997), whereas scant research has been performed on written data (Grafmiller, Szmrecsanyi & Hinrichs, 2018; Hinrichs, Szmrecsanyi & Bohmann, 2015).

For Chinese speakers of English, the process of using zero relative pronouns appropriately also involves different contributing factors. However, little attention is paid to the variation of relative pronoun omission among nonnative speakers and even less is given to written texts (Flanigan & Inal, 1996). Yet, previous research has found that zero relativizers were more favored in speaking over writing (Guy & Bayley, 1995), suggesting there were differences in uses between the two modalities. To fill the research lacuna, the present research focused on

variation of relative pronoun omission among Chinese speakers of English at the tertiary level in writing. Utilizing the contributing factors of relative pronoun omission found in previous research, the current study attempted to build a parsimonious model that could predict the use of relative pronoun omission through VARBRUL analysis.

The structure of the study is as follows. The following section reviews studies on relative pronoun omission and factors affecting its occurrences. The section after that describes the participants, data collection and statistical analysis. After that are the findings of VARBRUL analysis and explanations of the contributing factors of relative pronoun omission. Finally, the study concludes by giving some pedagogical implications.

#### Studies on Zero Relative Pronouns

The choice of relative pronouns has received a great deal of attention from variationists in past decades. A plethora of diachronic and synchronic research on the variation of relative pronouns indicates that linguistic and extra-linguistic factors have different effects on the use of relative pronouns (Guy & Bayley, 1995; Romaine, 1980; Tagliamonte, Smith & Lawrence, 2005).

An earlier study conducted by Romaine (1980), which examined the relativizers in written and spoken Scottish English from a historical linguistics perspective, found that there was a genre effect on relativizer use. The more stylistically and syntactically difficult the texts were, the more likely the use of *wh-* would occur in the Middle Scots language. On the contrary, in less formal texts, *that* and zero were found. The genre, or modality, effect was also found in the study of variation in American Standard English by Guy and Bayley (1995). They investigated the spoken data involving the choice of relative pronouns from The White House Transcripts and written data from a variety of academic articles, showing that the formality of both sources was different. Although the findings revealed that the zero form was the least frequently used relative pronoun, it was found to be more favored in informal conversation than in formal written texts. Furthermore, this study showed that zero relative pronouns were more favored by the non-subject position (That was the man  $\emptyset$  I saw yesterday), followed by animacy (He is *the man*  $\emptyset$  I know) and then adjacency to the antecedent (That was *the thing*  $\emptyset$  I did).

Flanigan and Inal (1996) used questionnaires and a sentence combining task to elicit how native speakers of English (NSs) and non-

native speakers (NNSs) of both longer and shorter lengths of residence in the U.S. used relative pronouns. In the sentence combining task, two sentences were combined into one by using a relative pronoun. The findings from the sentence combining tasks showed that NNSs only favored *wh-*, whereas NSs only favored zero. The results of the questionnaires showed that NSs preferred zero in speaking and *that* was preferred for writing. However, NNSs preferred to use *wh-* for writing, whereas zero pronouns were less often used. Regardless of the writer's nationality, overall zero relative pronouns were fairly uncommon in writing. On the other hand, zero relative pronouns were used more in speaking. The findings therefore indicated that there was a clear difference in the proportion of zero relative pronouns in writing and speaking.

In addition to the extra-linguistic factors, a great number of research studies have indicated that the choice of zero relativizers could be attributed to psycholinguistic and linguistic variables. For instance, Lehmann (2002) explored zero subject relativizers in spoken British and American corpora. The findings revealed that the British used more zero relative pronouns than the Americans did. Furthermore, matrix sentence constructions, such as existentials (*There is a book Ø she used to have*) and it-clefts (*It is the car Ø I used to drive*), were highly associated with subject zero relative pronouns.

Levey and Hill (2013) examined relativizer omission in Canadian spoken English. In terms of the distribution of the relative pronouns, the results revealed that *that* was the most prevalent one in Canadian English, followed by the zero form, *who* and *which*. In terms of linguistic factors, the study showed that matrix construction types, adjacency, unique antecedent NPs (as in “that was *the worst job* Ø I ever had”), the subject of relative clauses (that is, personal pronouns, as in “that was one of the things Ø *he* did when he was living elsewhere”), length, and the lexical specificity of head NPs (nonspecific NPs as in “it's *something* Ø I've thought about”) all favored the zero relativizer at the non-subject position. The researchers also found that the strength of the factors was matrix sentence constructions > adjacency > types of antecedent NPs > the subject of relative clauses > the length of relative clauses > the lexical specificity of head NPs.

Jaeger and Wasow (2006) investigated how the degree of accessibility influenced the occurrence of relativizers in non-subject relative clauses. The degree of accessibility here refers to the degree of ease for the interlocutors to construct a referent independent of conversations, which

are influenced by several factors such as number of antecedents or animacy. Using an informal spoken English corpus, the results showed that inherent accessibility factors such as the number of antecedents (*The things* Ø I've done vs. *The man* who called me), the referential use of pronouns (One thing Ø I do at school is playing basketball), and animacy resulted in relativizer variation. Specifically, singular referents and the referential use of pronouns favored the zero relativizer. Non-subject relative clauses with an expletive *it*- subject or an existential *there*- favored zero relative pronouns. Given the findings, the author hypothesized that the more accessible a referent was, the less frequently a relativizer would be used.

Hinrichs, Szmrecsanyi and Bohmann (2015) and Grafmiller, Szmrecsanyi and Hinrichs (2018) investigated the choices between *that/which* in subject relative clauses and the choices among *that/which/zero* in non-subject relative clauses. Using corpora from American and British standard written English, the findings showed that the length and the distance between a relative pronoun and an antecedent and lexical properties (such as nouns) could help predict relativizer omission. Specifically, the researchers discovered that the longer the length of an antecedent or a relative clause, the more likely a zero relative pronoun would not be selected. The results thus echo the Complexity Principle (Rohdenburg, 1996), which contends that when the language environment is more complex, more explicit forms will be preferred. Moreover, as in Guy and Bayley (1995), zero relativizers tended to appear in informal texts. Finally, it was found that the zero relativizer was preferred in American publications.

Tottie and Rey (1997) investigated spoken African American Vernacular English. Their data were taken from 10 speakers, and they found that there was a high tendency of using the zero relative pronoun in non-adverbial relative clauses. Also, the VARBRUL analysis indicated that the syntactic function of an antecedent head in a matrix clause, non-human antecedents, and adjacency favored zero relative pronouns. The hierarchy of the strength of the factors was the syntactic function of an antecedent > adjacency > animacy.

Bayley (1999) obtained data from interviews and investigated relative pronoun variation in Mexican-American English (MAE) among different social classes and ages. The results showed that MAE speakers used *that* the most, followed by zero and *wh*- forms. In addition, the non-subject position in the relative clause strongly favored zero, followed by

adjacency.

Tagliamonte, Smith and Lawrence (2005) explored zero relative pronouns in Northern British vernacular spoken English. The study revealed that the zero relative pronoun was the second most common relativizer. The researchers, from a diachronic perspective, explicated that such high frequency of zero relative pronouns was due to *wh*- relative pronouns being introduced to spoken English later than *that* and zero. Therefore, *wh*- relative pronouns did not penetrate completely into the local varieties. In terms of the linguistic constraints on zero relativizers, the results showed that sentence structures (existentials, as in *there are things Ø I don't know*, possessives, as in *I have a class Ø I hate*, and clefts, as in *it is the bag Ø she wants*), indefinite NPs and indefinite pronoun antecedents, and simple short relative clauses all favored zero subject relativizers. The hierarchy of the factors in Lowland Scottish English was sentence structures > the type of antecedents > the length of the relative clause. The hierarchy of the factors in Northern Irish English and Northwestern British English was sentence structures > the length of the relative clause. In addition, sentence structures (existentials and clefts), and simple short relative clauses favored object zero relative pronouns. The hierarchy of the factor groups was sentence structures > the length of relative clauses in Lowland Scottish and Northwestern British English.

Levey (2006) examined relativization strategies of spoken London vernacular English among adolescents of different genders. The findings demonstrated that London vernacular speakers preferred to use the relativizer *that* the most, while *wh*- relativizers were the least preferred. Also, the VARBRUL results showed that the zero form was strongly favored in the non-subject position, followed by a shorter relative clause (five words or less). However, genders were not a contributing factor.

Rickford (2011) examined zero relativizers in Guyanese, Jamaican, Appalachian, and African American Vernacular English (AAVE) spoken corpora. Through VARBRUL analysis, the findings showed several significant factors and the different hierarchy of the strength of the factors. In Guyanese English, the strength of zero relativizers at the subject position was adjacency > sentence structures (clefts, existentials and possessives). In Jamaican English, definite pronouns, superlative NPs, and indefinite pronouns had the most strength to predict zero relativizers at the subject position, followed by sentence structures (clefts and existentials). In terms of object relativizer omission, the results revealed certain types of antecedents (superlative NPs, definite pronouns and indefinite NPs)

## *RELATIVE PRONOUNS OMISSION VARIATION*

were the significant factor in Guyanese English. In Jamaican English, the hierarchy for object relativizer omission was the type of antecedents (superlative NPs, definite pronouns and indefinite pronouns) > sentence structures (clefts) > adjacency. In AAVE, the hierarchy was the type of antecedents (superlative NPs and indefinite pronouns) > adjacency > the humanness of antecedents (i.e., non-human). Based on the findings, the researcher concluded that the reasons for relativizer omission could be due to a variety of specific or universal language processing.

In conclusion, previous research on relative pronoun omission showed several contributing linguistic factor groups, including syntactic positions, animacy, adjacency, sentence structures, types of antecedent NPs, the length of relative clauses, the number of antecedents, and lexical properties. However, the strength of the factor groups varied from one study to another. Furthermore, researchers have discovered that relative pronoun omission was less frequent in writing than speaking, yet native speakers of English might show a higher rate of omission than nonnative speakers. Also, relative pronoun omission was found less frequently in less formal texts.

Despite the extensive research dedicated to this topic, very little attention has been paid to nonnative speakers of English, such as Chinese learners of English. Also, as variation among nonnative speakers has gradually been accepted as natural and legitimate (Davies, 1991; Dewey, 2012; Jenkins, 2006), studies focusing on this aspect can also shed light on language universals (Ranta, 2013). Research on relative pronoun omission in writing is still scanty; therefore, more studies can be conducted on written data. To close the research gap, the present study investigated how Chinese learners of English at the tertiary level employed zero relative pronouns in writing.

### **Research Questions**

Based on the abovementioned literature, the current research responded to the following questions.

1. What is the distribution of relative pronoun omission among Chinese learners of English at the tertiary level in their writing?
2. What are the factors that systematically influence the omission of relative pronouns in writing among Chinese learners of English at the tertiary level?

## **METHODOLOGY**

### **The Research Site and Participants**

The written data were collected from two courses that used English as a medium of instruction at a national university in Taiwan in fall semester, 2013 and 2014. The courses were offered by the General Education Center to students from all departments. There were 59 students in the class of 2013. Except for a Russian and a Chinese-American, the remaining 57 students were Taiwanese. As for the class of 2014, there were 46 Taiwanese along with two Chinese Malaysians, a Russian, two Hong Kongers and a Macanese. Although the classes did not have any specific English proficiency requirements for entrance, all of the students in the two classes reported that they had achieved a score of at least 760 on the Test of English for International Communication (TOEIC). Such a score is considered to be equivalent to high-intermediate level English proficiency. In considering that those expatriate students, such as those from America, Hong Kong or Malaysia, may be accustomed to using English as their medium language for life or early education, they were excluded from this study. The Russian students were also excluded to prevent the possibility of skewing the results. After excluding the written work from the non-Taiwanese students, a total of 396 written assignments were collected for analyses.

### **Data Collection**

In the two classes, students learned how linguistics could be applied in different fields, such as in forensic cases or cross-cultural communication. The students needed to submit a written assignment after a new lesson was taught every week. For each assignment, they were asked to summarize the material discussed during the week, identify the controversies in the content, compare and contrast their thoughts before and after the class, and provide critical analysis on the presented information. A total of 396 written assignments were collected for analyses. On average, the length of an assignment is 346 words and the average number of paragraphs is 3.9. The combined total number of words in all assignments is 123,760.

### **Data Analysis**



## RELATIVE PRONOUNS OMISSION VARIATION

### *Statistical method.*

Given that the choice of zero relativizers and the influential factors are categorical and discrete, the VARBRUL analysis is used for the statistical analysis. In order to uncover the variation rule of the omission of relative pronouns, the data were computed with the VARBRUL in GoldVarb 2001, an application of multivariate analysis specifically designed for Windows (Robinson, Lawrence & Tagliamonte, 2001). As a type of logistic regression, the VARBRUL analysis adopts a log-additive method and a logit link function for the combination of probabilities and calculating of a log-linear model (Paolillo, 2002). The model is calculated by using the equation of  $\log(1/1-p) = \text{input probability} + \text{sum of factor effects}$ . For detailed mathematical explanations of the model, refer to Paolillo (2002). The analysis enables researchers to examine the interactions among various factors and generate a parsimonious model for variation (Rand & Sankoff, 1990; Sankoff, 1988a). Its use has been attested to be informative in SLA variation research (Bayley, 1994, 1996; Young, 1991, 1996; Tarone, 1979, 1983, 2007).

Using the logarithmic link function, the VARBRUL is designed to change the components of the model into proportion-like probabilities (ranging from zero to one), which are called weights in this program. If the weight is above 0.5, it means the linguistic factor favors the variant while a value below 0.5 means the factor inhibits it. The VARBRUL also computes a step-wise model build and hypothesis test so researchers can easily generate a parsimonious model for a given variant. In other words, the VARBRUL performs a stepping up analysis, in which another variable is added whenever one variable is significant. Also, it performs a stepping down analysis in which all variables are initially grouped together and then one variable is removed at a time to see which one helps explain more variance. The step-wise analysis therefore generates the best combination of variables for a parsimonious model and marks the variables as significant. For a detailed procedure of using the VARBRUL analysis and explaining the results, see Young and Bayley (1996).

### *Coding.*

Given the focus of the study, the researcher and a native English-speaking coder first systematically extracted every restrictive relative clause from the essays as shown in (1). In (1), the relativizer who is restricted to denoting the specific group “people” instead of a group in

general. Following Levey and Hill (2013), several types of relative clauses were excluded from the analyses, including non-restrictive relative clauses, adverbial relative clauses (what, where, why), free or headless relative clauses, whiz-deletion, and resumptive pronoun surfaces in the relative clause. The reason non-restrictive relative clauses were excluded was they were restricted to wh- relative pronouns only (Ball, 1996). Adverbial relative clauses were also excluded because they were constrained by other linguistic factors and took different variants (what, where, why) other than non-adverbial relative clauses (which, who and that). Free or headless relative clauses were excluded because they did not have an antecedent and thus could not be coded. Whiz-deletion, the deletion of a wh-word plus a verb (usually a be verb) in a relative clause, was excluded because the rule required omitting an additional verb. The words “who are” in the relative clause “men (who are) sharpening knives are looking at us” can be deleted to form a whiz-deletion. Resumptive pronoun surface construction was not included because there was no way to identify whether the relative clause referred to its antecedent or the pronoun (for instance, That is the boy that he cries loudly). Also, direct quotations, parts that were detected as plagiarism and not-understandable sentences such as “However, some dangers or regardless which may cause a lot of troubles” were also excluded. After the extraction, a total of 416 restrictive relative clauses were found.

(1) Therefore, people [who] do not sign for the opt-in agreement are not always unwilling to donate their organs.

After extraction, the researcher commenced coding. First, the researcher and the coder, a native speaker of English with more than 10 years of experience of English teaching at the tertiary level, identified relative pronoun tokens. Since the purpose of the study explored the structured patterns of relative pronoun omission, this type of use was included. Tokens that could be seen as ungrammatical, such as the subject relativizer omission “I admire the people  $\emptyset$  have courage to speak out” were also included. As the present study adopted VARBRUL analysis (Sankoff, 1988b), the researcher first conducted an initial VARBRUL run<sup>1</sup>

---

<sup>1</sup> The total Chi-square value was 247.6345, which was above the critical value 29.588 (df = 12, p=.001). Since a logistic regression model requires factors to be independent from each other, the researcher combined definite and unique antecedents because they

## RELATIVE PRONOUNS OMISSION VARIATION

to test whether interactions existed among the factors (Paolillo, 2002). After the computation, the factors selected were used as the coding scheme in Table 1. The dependent variable includes the zero relative pronoun and other relative pronouns. The independent variables include sentence structures (existentials, *if*-clefts, possessives and others), syntactic functions (subject and non-subject), different types of antecedent NPs (definite NPs, indefinite NPs and pronominal NPs), as well as the length of relative clauses (short and long). Each factor was assigned a number according to the coding scheme. An example of the coding scheme can be seen as follows: “The Taiwanese chocolates do melt in your hands because of the heat, but it is the way [Ø] the propaganda is told.” was coded as 12212. The first number, 1, represents the zero relative pronoun. The second number, 2, represents an *it*-cleft. The third number, which is also 2, represents a non-subject position. The remaining numbers have their own assigned meanings.

The researcher coded all the tokens and the native speaker coder coded half of the tokens. The inter-coder reliability was 97%. The researcher and the coder discussed the disagreements and reached a consensus. Therefore, the remaining 3% of the tokens were recoded and included.

## RESULTS

Table 2 illustrates the overall distribution of the relative pronouns in the data. It is evident that the zero variant is the most commonly used form, followed by *who* and *that*. *Which*, on the other hand, is the least commonly used variant in the data.

The results of the subsequent VARBRUL procedure are shown in Table 3. The results show the total chi-square value is 20.2875 (df=6, p=.001), suggesting all the factors and factor groups are independent. In other words, the model is reliable. Table 3 shows which factor groups

---

inherently denoted definiteness. The researcher also combined definite and indefinite pronoun antecedents as they were all pronouns. Also, the researcher eliminated animacy, adjacency and number of the antecedents because they were not selected in the model during the process. The researcher also eliminated the subject of the relative clause because pronouns as subjects might overlap with the factor of a pronominal antecedent under types of antecedents.

contribute significant effects to the zero variant choice. In Table 3, we can see tokens of zero relativizers and their percentages in each factor. We can also see probability weights, which refer to the likelihood that a factor can predict the use of zero relativizers. The value is between 1 and 0. A value above 0.5 means the factor promotes the occurrence of zero relative pronouns, whereas a value below 0.5 means it inhibits the occurrence. The higher the value, the stronger the factor can predict its occurrence. Also, the range refers to the relative strength of a factor group in comparison with other factor groups. The range is calculated by subtracting the lowest factor weight from the highest factor weight in a factor group. The higher the value of the range, the stronger the factor group could predict the occurrence of the zero relativizer. Finally, the input probability is the overall indication of the strength of the rule. It shows the likelihood that a target variant will appear. The input probability is 30.8%, which indicates that there is a 30.8% likelihood that the relative pronoun is zero for the participants. Furthermore, the findings show that sentence structures, syntactic functions, the type of antecedents, and the length of relative clauses can help one to predict the choice of zero relative pronouns given that the factor weights are significant.

When considering sentence structures, existentials strongly favor zero relative pronouns. Possessives also favor zero relativizers but the tendency is not as strong. Conversely, other constructions and clefts disfavor zero relativizers. With regard to syntactic functions, the non-subject position strongly favors zero relativizers while the subject position disfavors them. As for the type of antecedent NPs, only definite antecedent NPs favor zero variants, but indefinite NPs and pronominal NPs disfavor them. In terms of the length of a relative clause, if the length is short, it favors omission of relativizers. On the contrary, long relative clauses disfavor it.

## **DISCUSSION**

The first research question is to examine the distribution of zero relativizers among Chinese learners of English in writing. Different from earlier research in which writing preferred *wh*- forms because it was a genre that required formality (Guy & Bayley, 1995), it was found that the zero variant was the most preferred form among the participants. The reason may be because the participants are nonnative speakers and the system of using zero relativizers is different in their L2. Therefore, the distribution does not resemble that of native speakers.

#### *RELATIVE PRONOUNS OMISSION VARIATION*

In addition to the overall frequency of the variants, the present study also demonstrated that the choice of zero relativizers had a systematic pattern. Moreover, the omission of relativizers is constrained by various linguistic factors, including sentence structures, syntactic functions, the type of antecedent NPs and the length of a relative clause. Based on the values of ranges, the different factor groups present a hierarchy of strength. That is sentence structures > syntactic functions > the type of antecedent NPs > the length of a relative clause. With respect to sentence structures, as Tagliamonte, Smith and Lawrence (2005) stated, sentence structures were the most attested factor that could predict relative pronoun omission. This statement also holds true in nonnative speakers' writing.

Specifically, the results showed that existentials and possessives favored relative pronoun omission but clefts did not. The reason these sentence structures were more likely to predict relative pronoun omission could be explained by the Predictability Hypothesis (Wasow, Jaeger & Orr, 2011), which states that in environments where a non-subject relative clause is more predictable, relativizers are less frequent. It should be noted that although Wasow, Jaeger and Orr did not develop a precise definition and the hypothesis was a preliminary one, they revealed a clear relationship between the occurrence of linguistic features and the use of relative pronouns. In other words, the hypothesis states that when certain linguistic factors appear, there will be a strong prediction that a non-subject relative clause will also appear. Under these circumstances where a non-subject relative clause is more predictable, a relativizer tends to be absent. Existentials and possessives are common structures co-occurring with non-subject relative clauses. As the two sentence structures are closely related to non-subject relative clauses, the use of relative pronoun omission is more predictable. Therefore, relative pronouns tend to be absent.

In terms of syntactic functions, the findings indicated that the non-subject position strongly favored zero relative pronouns. The reason the non-subject position was the norm may be because of the frequency effect. Reali and Christiansen (2007) mentioned that the representations of structures of relative clauses could be shaped by language use and frequency of occurrence. Higher frequency occurrences would lead to stronger representation of a structure. As many studies on native English varieties have shown, zero relativizers were strongly favored at the non-subject position (Bayley, 1999; Levey, 2006; Levey & Hill, 2013; Tagliamonte, Smith & Lawrence, 2005). Therefore, from a usage-based

perspective, it can be observed that the participants were influenced by the frequency of relativizer omission and its strong preferences at the non-subject position. In terms of the reason the non-subject position favored the zero variant, it could be because of ambiguity avoidance (Temparley, 2003). If we compare relative pronouns that associate with the subject position, as in (2) and (3), with relative pronouns that associate with the non-subject position, as in (4) and (5), it is obvious that (2) and (3) are more ambiguous than (4) and (5). Sentence (2), in particular, when the zero variant is used at the subject position, it will cause a garden path effect and thus result in processing difficulty. Accordingly, the zero variant is more prominent at the non-subject position.

- (2) The man hired me was very stingy.
- (3) The man who hired me was very stingy.
- (4) The man I hired was very stingy.
- (5) The man whom I hired was very stingy.

The findings also found two relatively weaker factors that favored the zero variant, that is, the type of antecedents and the length of relative clauses. As for the types of antecedents, from the perspective of lexical inherent accessibility, Jaeger and Wasow (2006) proposed that less accessible referents required a relativizer more frequently. Here, the accessibility means “the ease with which the mental representation of some potential referent can be activated in or retrieved from memory” (Bock & Warren, 1985:50). Therefore, compared with indefinite NPs which take more effort to retrieve from memory, definite NPs are more accessible or easier to be retrieved from memory. Hence, a relative pronoun is more likely to be omitted when an antecedent NP is definite. Moreover, Fox and Thompson (2007) found that the unique head NP was one of the characteristics that correlated with the absence of relativizers. Speakers tended to treat this type of sentence construction as a monoclausal combination as well as a single processing unit. In the initial VARBRUL run, the unique head NP factor was combined with definiteness because they were fundamentally similar. Therefore, in analogy, the reason definiteness favored the zero variant might be attributed to the same explanation. Lastly, based on the Predictability

## *RELATIVE PRONOUNS OMISSION VARIATION*

Hypothesis (Wasow, Jaeger & Orr, 2011), determiners increased the possibility that a non-subject relative clause will appear; therefore, a relativizer is less needed when a definite antecedent exists.

Finally, the length of relative clauses also predicted the zero relative pronoun omission. Specifically, short relative clauses favored zero relativizers. This finding is in accordance with Hinrichs, Szmrecsanyi and Bohmann (2015). Fox and Thompson (2007) proposed the idea of monoclausality to explain the correlation between certain linguistic features such as shorter relative clauses and the absence of relative pronouns. Fox and Thompson stated that the idea of monoclausality was a continuum where certain linguistic features co-occur with zero relative pronouns and thus make two clauses more monoclausal. Those linguistic features therefore fall at the end of monoclausality. Fox and Thompson also explained that the reason shorter relative clauses fall at the end of monoclausality was due to the frequency of use. In other words, shorter relative clauses were used more frequently with a zero relativizer. Such frequency influences the way speakers form relative clauses. Therefore, participants can be influenced by the high frequency of the pattern where zero relativizers normally co-occur with short relative clauses. Also, from the psycholinguistic point of view, higher frequency usages are easier to process or access from memories (MacDonald, 2013). Fox and Thompson then stated that the pattern where shorter relative clauses co-occur with zero relativizers could be easier stored by its users because of its high frequency and therefore also tends to be often reproduced.

In this study, the existence of zero relativizers revealed a systematic pattern of variation in writing among Chinese learners of English. Furthermore, most of the factors of predicting relative pronoun omission can be explained by language processing, predictability or the frequency effect. Although it appears that the reasons for variation are associated with the intralingual effect, it should be noted here that the participants could have been affected by other potential factors such as their L1 because the formation of relative clauses in Chinese is different from their formation in English. However, comparing the effects from the intralingual perspective is not the focus of this study. That investigation will be left to future research.

### **CONCLUSION**

The study investigated the distribution of zero relativizers and the

variable rule of this linguistic feature among Chinese learners of English at the tertiary level in writing. The findings indicated that the zero variant was the most frequent form used by the participants. This may be due to the fact that nonnative speakers have different preferences from native speakers. In addition, the choice of zero relative pronouns was systematically rule-governed by linguistic constraints. That is, the omission of a relative pronoun most frequently occurred when a sentence was an existential or possessive sentence with a definite NP as an antecedent. Moreover, the zero relativizer was favored in the non-subject position and when the following relative clause was short.

The systematic occurrence of zero relativizers has indicated that, similar to native speakers of English, Chinese learners of English are also constrained by linguistic and psycholinguistic factors such as language processing, predictability and the frequency effect. Therefore, the choice of zero relativizers among nonnative speakers is bolstered with reasonable explanations. Based on the findings, some implications can be deduced. First, the results indicated that Chinese learners of English can be seen as legitimate and competent English users given that they are influenced by the same factors as native English speakers are. Second, zero relativizers receive relatively less attention than *wh*- relative pronouns in English teaching and learning. The rules of omitting relative pronouns can also be unclear for learners. Therefore, for language teachers and learners, the results present the rule of using zero relativizers among proficient English users. Teachers and learners can make use of the hierarchy of linguistic constraints to understand that certain sentence structures, non-subject position and definite antecedents are very likely associated with the relativizer omission.



## RELATIVE PRONOUNS OMISSION VARIATION

### REFERENCES

- Ball, C. N. (1994). Relative pronouns in *it*-clefts: The last seven centuries. *Language Variation and Change*, 6(2), 179–200.
- Ball, C. N. (1996). A diachronic study of relative markers in spoken and written English. *Language Variation and Change*, 8(2), 227–258.
- Bayley, R. J. (1994). Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English. In S. Gass, A. Cohen, & E. Tarone (Eds.) *Research methodology in second-language acquisition* (pp.157–181). Hillsdale, NJ: Erlbaum.
- Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley & D. Preston (Eds.), *Second language acquisition and linguistic variation* (pp.97-120). Amsterdam: John Benjamins.
- Bayley, R. (1999). Relativization strategies in Mexican-American English. *American Speech*, 74(2), 115-139.
- Bayley, R. (2000). Second-language acquisition and variationist linguistics. *American Speech*, 75(3), 288-290.
- Bock, J. K. & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47-67.
- Davies, A. (1991). The notion of the native speaker. *Journal of Intercultural Studies*, 12(2), 35-45.
- Dewey, M. (2012). Towards a post-normative approach: Learning the pedagogy of ELF. *Journal of English as a Lingua Franca*, 1(1), 141-170.
- Flanigan, B. O. & Inal, E. (1996). Object relative pronoun use in native and non-native English: A variable rule analysis. *Language Variation and Change*, 8(2), 203-226.
- Fox, B. A. & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 66(2), 297-316.
- Fox, B. A. & Thompson, S. A. (2007). Relative clauses in English conversation: Relativizers, frequency, and the notion of construction. *Studies in Language*, 31(2), 293-326.
- Grafmiller, J., Szmrecsanyi, B. & Hinrichs, L. (2018). Restricting the restrictive relativizer. *Corpus Linguistics and Linguistic Theory*, 14(2), 309-355.
- Guy, G. R. & Bayley, R. (1995). On the choice of relative pronouns in English. *American Speech*, 70(2), 148-162.
- Hinrichs, L., Szmrecsanyi, B. & Bohmann, A. (2015). *Which-* hunting and the standard English relative clause. *Language*, 91(4), 806-836.
- Jaeger, T. F. & Wasow, T. (2006). Processing as a source of accessibility effects on variation. In R. T. Cover & Y. Kim (Eds.), *Proceedings of the 31<sup>st</sup> annual meeting of the Berkeley Linguistic Society* (pp. 169-180). Ann Arbor, MI: Sheridan Books.
- Jenkins, J. (2006). Points of view and blind spots: ELF and SLA. *International Journal of Applied Linguistics*, 16(2), 137-162.
- Lehmann, H. (2002). Zero subject relative constructions in American and British English. In P. Peters, P. Collins & A. Smith (Eds.), *New frontiers in corpus research* (pp. 163-177). Amsterdam: Rodopi.

- Levey, S. (2006). Visiting London relatives. *English World-Wide*, 27(1), 45-70.
- Levey, S., & Hill, C. (2013). Social and linguistic constraints on relativizer omission in Canadian English. *American Speech*, 88(1), 32- 62.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontier in Psychology*, 4, 1-16.
- Paolillo, J. (2002). *Analyzing linguistic variation: Statistical models and methods*. California: CSLI Publication.
- Rand, D. & Sankoff, D. (1990). *GoldVarb: A variable rule application for the Macintosh*. Montreal, Canada: Centre de recherches mathématiques, Université de Montréal.
- Ranta, E. (2013). *Universals in a universal language? Exploring verb-syntactic features in English as a lingua franca* (Unpublished doctoral dissertation). University of Tampere, Tampere.
- Reali, F. & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1-23.
- Rickford, J. R. (2011). Relativizer omission in Anglophone Caribbean Creoles, Appalachian, and African American Vernacular English [AAVE], and its theoretical implications. In E. M. Bender & J. Arnold (Eds.), *Language from a cognitive perspective: Grammar, Usage, and Processing* (pp.139-160). CA: CSLI Publications.
- Robinson, J., Lawrence, H. & Tagliamonte, S. (2001). GOLDVARB 2001: A multivariate analysis application for Windows. Retrieved from [http://www.romanistik.uni-freiburg.de/pusch/Download/variacionismo/GoldVarb2001\\_User\\_manual.pdf](http://www.romanistik.uni-freiburg.de/pusch/Download/variacionismo/GoldVarb2001_User_manual.pdf)
- Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics*, 7(2), 149-182.
- Romaine, S. (1980). The relative clause marker in Scots English: Diffusion, complexity, and style as dimensions of syntactic change. *Language in Society*, 9(2), 221-247.
- Sankoff, D. (1988a). Sociolinguistics and syntactic variation. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey. Language: The socio-cultural context* (Vol.4) (pp. 140- 161). Cambridge: Cambridge University Press.
- Sankoff, D. (1988b). Variable rules. In U. Ammon, N. Dittmar & K. J. Mattheier (Eds.), *Sociolinguistics: An international handbook of the science of language and society* (Vol. 2) (pp. 984-997). Berlin: Walter de Gruyter.
- Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
- Tagliamonte, S., Smith, J. & Lawrence, H. (2005). No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change*, 17(1), 75-112.
- Tarone, E. (1979). Interlanguage as chameleon. *Language Learning*, 29(1), 181-191.
- Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, 4(2), 142-164.
- Tarone, E. (2007). Sociolinguistic approaches to second language acquisition research—1997-2007. *Modern Language Journal*, 91(1), 837-848.
- Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language*, 79(3), 464-484.

*RELATIVE PRONOUNS OMISSION VARIATION*

- Tottie, G. & Rey, M. (1997). Relativization strategies in earlier African American Vernacular English. *Language Variation and Change*, 9(2), 219-247.
- Wasow, T., Jaeger, T. F. & Orr, D. M. (2011). Lexical variation in relativizer frequency. In H. J. Simon & H. Wiese (Eds.), *Expecting the unexpected: Exceptions in grammar* (pp. 175-196). NY: De Gruyter Mouton.
- Young, R. (1991). *Variation in interlanguage morphology*. NY: Peter Lang.
- Young, R. (1996). From-function relations in articles in English interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp.135-176). Philadelphia: John Benjamins.
- Young, R. & Bayley, R. (1996). VARBRUL analysis for second language acquisition research. In R. Bayley & D. R. Preston. (Eds.), *Second language acquisition and linguistic variation* (pp. 253–306). Philadelphia: John Benjamins.

*Hsuan Yu Tai*

**ACKNOWLEDGMENT**

The author would like to thank the anonymous reviewers of the *Taiwan Journal of TESOL* for their constructive comments to help improve earlier versions of this paper. The author's sincere gratitude also goes to Prof. Victoria Rau for her thoughts on the research.

**CORRESPONDENCE**

*Hsuan-Yu Tai, Language Center, National Chin-Yi University of Technology, Taichung, Taiwan*  
*Email address: tsuyoshi@ncut.edu.tw*

**PUBLISHING RECORD**

Manuscript received: April 22, 2019; Revision received: September 28, 2019; Manuscript accepted: October 22, 2019

APPENDIX

**Table 1. Dependent and Independent Variables of VARBRUL Analysis**

---

Dependent variable
1. zero relative pronoun      2. others

---

Independent variables

a. Sentence structures

1. existentials

Ex: There are several common misunderstandings  $\emptyset$  people think about voice-print ID system.

2. it- clefts

Ex: The Taiwanese chocolates do melt in your hands because of the heat, but it is the way  $\emptyset$  the propaganda is told.

3. possessives

Ex: I think McDonald's has the right to defend the reputation  $\emptyset$  they have built up.

4. others

Ex: People should not using words that have similar pronunciation in any serious report.

b. Syntactic function of the relativizer

1. subject

Ex: And there's no one  $\emptyset$  can perfectly imitating a phonological system.

2. non-subject

Ex: When the soldier that I control have to confirm a thing to his headquarters, they only use the word positive and negative but never using reactive.

c. The types of antecedents

1. definite antecedent

Ex: When the soldier that I control have to confirm a thing to his headquarters, they only use the word positive and negative but never using reactive.

---

2. indefinite antecedent

Ex: When we're walking on the street, we can see various names Ø represent different kinds of stores.

3. pronominal NPs

Ex: There are several reasons for those who are not willing to donate their organs.

c. Length of the relative clause

1. short (three words or fewer)

Ex: After all, everyone has the right to use whatever word Ø they want.

2. long (more than three words)

Ex: Still, there are several common misunderstandings Ø related to voice-print ID system.

---

*RELATIVE PRONOUNS OMISSION VARIATION*

**Table 2. Overall Distribution of Relative Pronouns**

Relativizers	Token	Percentage
Zero	128	30.7%
Who	119	28.6%
That	96	23%
Which	73	17.5%
Total	416	100%

**Table 3. Variable Rule Analysis of Factors Contributing to the Probability of Zero Relative Pronouns**

Factor groups	Use of zero relative pronoun
Sentence structure	Token/Total (%)
	Probability Weight
Existentials	29/ 53 (54.1%)
	0.944*
Possessives	22/ 55(40%)
	0.637*
Others	71/ 268 (26.4%)
	0.434*
Clefts	6/40 (14.3%)
	0.212*
Range	0.732
Syntactic function	Token/Total (%)
	Probability Weight
Subject	21/268 (7.8%)
	0.216*
Non-subject	107/148 (72.3%)
	0.912*
Range	0.696
Type of antecedent NP	Token/Total (%)
	Probability Weight



*RELATIVE PRONOUNS OMISSION VARIATION*

Definite	78/186 (41.9%)
	0.641*
Indefinite	36/ 195(18.5%)
	0.401*
Pronoun	14/35(4%)
	0.299*
Range	0.342
<hr/>	
Length of relative clause	Token/Total (%)
	Probability Weight
Short	53/105 (50.5%)
	0.659*
Long	75/ 311 (24.1%)
	0.445*
Range	0.214
<hr/>	
Total	128/416 (30.8%)
Input probability (corrected mean)	0.308 (30.8%)
Application value	1
Total Chi-square = 20.2875 < 22.458	
(df = 6), p=.001	
Chi-square/cell = 0.7514	
Log likelihood = -136.409	
<hr/>	