The Application of Corpora in Teaching Grammar: The Case of English Relative Clause

Rahman Sahragard, Ali Kushki and Ehsan Ansaripour*
Shiraz University


The study was conducted to see if the provision of implementing corpora on English relative clauses would prove useful for Iranian EFL learners or not. Two writing classes were held for the participants of intermediate level. A record of 15 writing samples produced by each participant was kept in the form of a portfolio. Participants’ portfolios in both the experimental and the control groups were analyzed to spot patterned errors. Having diagnosed the errors, both groups were instructed on the use of English relative clauses. The participants in the experimental group were instructed using the corpora as printed materials. The control group, however, were instructed using explicit definition, discussion and exemplification. Then, both group participants’ portfolios were returned and they were asked to self correct their misapplied relative clauses. Interestingly, both groups had improved significantly. Two Chi-square tests on the use of *that* and *which* were run before and after classroom procedures on both groups. The first one suggested similar performance for the two groups but the second one favored the experimental group. The study suggests that applying corpora is an effective way to make students aware of their errors which ultimately leads them to self correction.

**Key Words:** relative clause types, Corpus Linguistics, relativizers, antecedents, consciousness-raising, noticing

1 Introduction

One of the most difficult syntactic structures to master is relative clauses. It has been understood that in the difficulty hierarchy they are ranked among the most difficult elements to learn even for native speakers (Ozeki & Shirai, 2007; Yabuki-Soh, 2007). Thus, investigating relative clauses doubles in importance for language learners from a different language background.
Relative clause research has enjoyed the attention of a number of SLA researchers (Ellis, 2008). However, most of the studies on relative clauses have been typological in nature. So, the purpose has been to understand how relative clauses work across languages and how they are different from and similar to one another. This bulk of study has resulted in a number of hypotheses like New Phrase Accessibility Hypothesis (NPAH), the Perceptual Difficulty Hypothesis (PDH), and the Structural Distance Hypothesis (SDH), and the Subject Object Hierarchy Hypothesis (SOHH) (Ellis, 2008; Izumi, 2003). Ellis (2008) summarizes the findings of 15 studies investigating L2 acquisition of relative clauses (Tarallo & Myhill, 1993; Gass, 1980; Hyltenstam, 1984; Pavesi, 1986; Eckman et al., 1988; Hawkins, 1989; Doughty, 1991; Wolfe-Quintero, 1992; Hamilton, 1994, 1995; Izumi, 2003; Matthews & Yip, 2003; O’Grady et al., 2003; Ozeki & Shirai, 2007; Yabuki-Soh, 2007). Of course, these studies include other languages with English receiving the predominant focus. However, as Ellis (2008) puts it, the bulk of the studies have a quasi-experimental design meaning that these studies do not consider any treatments and have been mainly focused on the analysis of non-experimental data. This implies that the treatment dimension of these studies is unspecified.

In addition, Ellis (2008) argues that all studies carried out so far have collected their data by resorting to controlled elicitation instruments like sentence joining tasks, picture description tasks, and grammaticality judgment tasks. The present study is different from the backdrop in three respects. First, the study is not just a diagnostic and descriptive one; rather it includes a treatment dimension trying to provide a condition for the learners to outflank the difficulty they have with producing English relative clauses. Second, the data was collected from the writing samples of 15 language learners taking writing lessons without employing any elicitation techniques. Hence, the authenticity and naturalness of the collected data is well taken into account in this study. Third, this is the first study, to our knowledge, that contrasts Persian and English relative clauses and proposes a treatment for the problems identified using corpora as materials. An exception is Yarmohammadi and Rashidi’s study (2009) which intuitively contrasts some selected structures of Persian and English one of which is relative clauses. However, their study does not propose any solution for the probable difficulty Iranian learners of English may have when learning English relative clauses.

1.1 Relative clauses in English and Persian

Swan (2005, p. 477) defines relative clauses as clauses beginning with question words (e.g. who, which, where) which are often used to modify nouns and some pronouns to identify people and things, or to give more information about them. These relative pronouns are called so because they
The Application of Corpora in Teaching Grammar:
The Case of English Relative Clause

play two roles in relative clauses. They play the role of a connective which connects two separate sentences. The second role they fulfill is that they function as pronouns since they substitute the noun as well (Yarmohammadi & Rashidi, 2009). In English choosing the appropriate relativizers is a function of its antecedent.

However, the case is different in Persian meaning that Persian relativizers are not antecedent-sensitive. Yarmohammadi and Rashidi (2009) maintain that in Persian there is only one connective /ke/ (that) which is used to connect dependant and independent clauses in a complex sentence and to provide additional information while in English there are words like who, which, who, etc. In other words, in Persian, /ke/ is used in cases for which in English we have who, whom, where, when, that. Nevertheless, in many cases, both which and that would be English equivalents of /ke/; so most Iranian learners of English tend to use them interchangeably and do not apply other available alternatives. Yarmohammadi and Rashidi (2009) assert that /ke/ acts only as a connective in Persian and the pronoun element which is existent in English is missing in Persian. In other words, the pronoun function of relative clauses in English is nonexistent in the Persian relative clause. This creates problems for Iranian language learners when trying to produce relative clauses in English. What this study aimed at fulfilling was to contribute the participants of the study to overcome the problems they had with learning the use of English relative clauses by drawing on Corpus Linguistics.

1.2 Corpus Linguistics

Computer technology has changed the face of our planet. This has dramatically gained momentum at the turn of 21st century. Now, with the unfailing obedience and availability of computers, different fields of study are increasingly flourishing. One of the beneficiaries of this technology boon has been language-related studies. The application and contribution of computers to language studies has come to be named “Corpus Linguistics (CL).” Reppen and Simpson (2002) assert that CL as an area of study has gained more popularity lately. Before going further, it is appropriate to start with what CL is. Since the literature abounds with definitions which only differ in their wording, our purpose here is not to provide an exhaustive list. Rather, attempt is made to just take account of the main components of corpus-based analyses of language and leave the defining shades of meaning to the individuals interested. Perhaps, the most comprehensive characterization of CL studies has been the one proposed by Reppen and Simpson (2002). They count four typical features of corpus studies of language as having an empiricist nature, having a corpus basis, using computers for analysis, and drawing upon both quantitative and qualitative analytical techniques. With these features in mind, the problem of differently
put definitions is militated against greatly. From its initiation into language studies, CL has widened our horizon of how language works. For many years, linguistics has been the arena of Chomskian linguistics which has been based on introspective analyses of native speakers as the way to anatomize language and language use was not enjoying its right place in language studies. Tognini-Bonelli (2001) maintains that the bulk of linguistic research requires evidence of language in use and corpus caters for such evidence. Now, thanks to this approach, linguistic patterns that could not be achieved before can be extracted. Cook (2001) observes that a particularly significant advance in the investigation of language use has been corpus linguistics. He adds that the approach makes it possible to search within seconds millions of actual language in use to gain information on the frequencies and combinations of intended words. Stubbs (2006) supports the essentiality of Corpus Linguistics for describing language use and believes that corpora show how lexis, grammar, and semantics interact. Also, Gavioli and Aston (2001) observe that corpora offer evidence of linguistic performance that can be doubtlessly helpful in deciding what to teach.

2 Method

Driven by the problems that participants of the study had with learning English relative clauses, the researchers aimed at troubleshooting the difficulty by resorting to relative clause corpora.

The present study was conducted in Chekad Language Institute, Shiraz, Iran. The institute is computerized with Internet connectivity. It provides on-line language tests and most language learners attending it are prospective IELTS exam takers. Also, all through courses, use is made of these modern equipment and some materials are made available online. Every language learner has unshared access to one computer during class time. The institute uses a net supporting software allowing teachers to control their students simultaneously. This is very helpful when it comes to writing classes where students are supposed to write down on a given topic in a word document environment. On the whole, the context features a multimodal one and meets the criteria of a multimodal world.

2.1 Participants

The participants of the study were 40 intermediate language learners aged 18 to 27. Although they had taken English classes, no one had attended exclusive writing classes before. Their only writing experience was their composition tasks they were required to fulfill as class homework. These tasks were limited in the sense that they were mostly at the level of sentence rather than suprasentence. For example, most of the participants said that they
only put simple sentences together to make their sentences look like a piece of composition.

2.2 Materials

Materials of the study were collected from eight writing classes taught by the researchers. Fifteen writing samples for each participant were kept in the form of a personal portfolio within 15 weeks. In the aggregate, 600 writing samples from the participants of the study were used in this study.

2.3 Data collection procedure

The data for the study comes from eight writing classes taught by the researchers. Each class consisted of 10 students. The classroom in which these writing classes were being held was equipped with ten computers, making it possible for each student to work with one. First, writing lessons were given by the researchers and then, language learners were asked to write on a topic based on the model practiced during instruction. A record of the writing samples produced by each participant was kept in the form of a portfolio. Writing sessions were held once a week. In the aggregate, 15 samples for each participant were gleaned. All the writing samples were numbered and dated to show their sequence. It was through the analysis of these portfolios that the researchers spotted the patterned and erroneous applications of relative clauses. After this recognition, three 30 minute treatment sessions were held to address the problem. There is nothing magical in this number three; it was only due to practical issues that the treatment was limited to these three sessions. For the experimental group, the learners were given English relative clause corpora (Corpus Concordance English) using online sources and printed ones in cases when internet connection was cut off. Then, they were asked to study them to induce regularities, here matching relativizers to their antecedents and inducing how their insertion was a function of their antecedents. All relativizers and their antecedents in the corpus were emboldened. Since the participants of the study were intermediate language learners, the researchers provided them with Persian translation of the words in the corpora to which they were not conversant. The idea was to get learners induce regularities in the corpora. In cases where they failed to do so, they were provided with deductive instruction. For the control group this period was spent nearly the same as above with the exception of corpus data in print-outs and on the internet. These sessions were run providing the different forms of relativizers in single sentences together with their definition and a question and answer discussion between the teacher and the participants. After the sessions for both groups, participants’ portfolios in the form of word document print-outs were given back to them, and were asked to make modifications in the relative clauses.
they had made use of in their writing samples in view of the treatment they received. To make certain that modifications are made by the participants, they were asked to fulfill the task in class during the three sessions without any conferring.

2.4 Data analysis

Having collected the data, they were analyzed to spot patterned errors. The only patterned error noticed in the data was the misapplication of relative clauses. First, the number and percentage of the relative clause errors made by each participant were counted. Then, after running the sessions, again the number and percentage of errors made by the participants were counted. In order to find out about any significant differences between the data before and after the sessions and between groups, three Chi-square tests were run using SPSS version 16. Having run the Chi-square tests for independence, the results of the tests were checked to see whether the assumptions concerning the ‘minimum expected cell frequency’ (Pallant, 2007) have not been violated. One more note should be added here. Since before the sessions, frequencies of all relative clause types included in this study except that and which were zero, the Chi-square test was run only for the that and which since for all the other relative clauses the pre-treatment frequency was 0 and so, automatically a significant difference was expected.

3 Results

In the following section, Figures 1 and 2 graphically show the frequency counts of relative clause types found in 600 writing samples of the participants of the study respectively before and after the class procedures. As Figure 1 represents, the participants of the study used only two relative clause types, namely that and which in their writing samples although this should not have been the case, given the antecedents of the relativizers used in the writing samples. This proves that they were either totally ignorant of other alternative relative clauses in English or were not sure of their use. Figure 1 shows the frequencies of the relative clause types used by both groups before doing any treatment to the groups.
The Application of Corpora in Teaching Grammar:  
The Case of English Relative Clause

Figure 1. Frequency of use for both the experimental and control groups before class procedures

The next step was to run a Chi-square test to see if there is any difference between the two groups in terms of their performance on ‘that’ and ‘which’. The results can be seen in Table 1 below:

Table 1. Chi-square for Both Groups on ‘that’ and ‘which’ before the Treatment

<table>
<thead>
<tr>
<th>Chi-Square Tests</th>
<th>Value df</th>
<th>As. Sig. (2-sided)</th>
<th>Ex Sig. (2-sided)</th>
<th>Ex Sig. (1-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>.677a</td>
<td>1</td>
<td>.411</td>
<td></td>
</tr>
<tr>
<td>Continuity Correctionb</td>
<td>.564</td>
<td>1</td>
<td>.453</td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>.677</td>
<td>1</td>
<td>.411</td>
<td></td>
</tr>
<tr>
<td>Fisher's Exact Test</td>
<td></td>
<td></td>
<td>.430</td>
<td>.226</td>
</tr>
<tr>
<td>Linear-by-Linear Association</td>
<td>.676</td>
<td>1</td>
<td>.411</td>
<td></td>
</tr>
<tr>
<td>N of Valid Casesb</td>
<td>982</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 132.73.

b. Computed only for a 2x2 table
The Chi-square test returned no significance difference between the two groups on the use of ‘that’ and ‘which’ ($X^2 = .677$, $p < .05$). This shows that the two groups nearly performed similarly on the use of ‘that’ and ‘which’ and the difference in frequency of errors is just due to chance factors.

Figure 2 tabulates the relative clause types used by the participants of the study after going through the procedures. As it is clear from the Figure, there is a reduction in the number of the two relative clause types exclusively used before the study and simultaneously an increase in the frequencies of other relative clause types.

![Figure 2. Frequency of relative clause use after the treatment on both groups](image)

As it is shown in Figure 1, the frequency counts for whom, when, where, and who were zero. But according to Figure 2, here the frequency counts for the same relative clause types increase up to 16, 25, 40, and 54 for the control group and 57, 84, 74, and 86 for the experimental group, respectively. The important point shown in the Figure is the simultaneous decrease in the frequency counts of the only two relative clause types used before the treatment. The increase in the other types can be justified by the fact that the participants of the study were not aware of the other alternatives for that and which that had most frequency counts before the study. Obviously any statistical comparison using Chi-square tests between the two groups is not necessary here as there are clear indications of the difference between the first performance of the groups and their second one. This means that the two groups have developed their knowledge as the result of classroom procedures.

To find out whether there was a statistically significant difference between the frequencies of the relative clauses of that and which used after the treatment for the two groups, another Chi-square test was run. Table 2 shows the results.
Table 2. Chi-Square for Both Groups on ‘that’ and ‘which’ after the Treatment

<table>
<thead>
<tr>
<th>Chi-Square Tests</th>
<th>Value</th>
<th>df</th>
<th>As. Sig. (2-sided)</th>
<th>Ex. Sig. (2-sided)</th>
<th>Ex. Sig. (1-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>6.450</td>
<td>1</td>
<td>.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuity Correction(^b)</td>
<td>5.995</td>
<td>1</td>
<td>.014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>6.522</td>
<td>1</td>
<td>.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fisher’s Exact Test</td>
<td></td>
<td></td>
<td></td>
<td>.014</td>
<td>.007</td>
</tr>
<tr>
<td>Linear-by-Linear Association</td>
<td>6.438</td>
<td>1</td>
<td>.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N of Valid Cases(^b)</td>
<td>546</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 78.92.

b. Computed only for a 2x2 table

As Table 2 indicates, the result of the Chi-square test shows a significant difference ($X^2 = 6.450$, $p < .05$) between the two groups in terms of their performance on *that* and *which* relative clauses. This means that the control group who went through a different instruction still used more of the two relative clauses than the experimental group. In other words, the experimental group has used other types of relative clauses in their second performance. This might be the result of their awareness gained through the procedures in the classroom.

**4 Discussion**

As referred to previously, the treatment was in the form of a relative clause corpus with its relativizers and their antecedents emboldened. The statistical difference resulting from the treatment can be justified and discussed from a couple of theoretical perspectives on SLA.

Fotos (1993) gives a succinct summary of the works carried out by Schmidt (1990) and Ellis (1990) who bolster viewing formal instruction as a consciousness-raising undertaking (Rutherford & Sharwood Smith, 1985). Based on Ellis’s (1999) view of language acquisition, it is through formal instruction that making learners aware of specific features of a target language and forming explicit representations of these features become possible. Hence, he advocates the implementation of Consciousness-Raising (CR) tasks. Ellis (1997, P. 160) defines a grammar consciousness-raising task as an “activity where the learners are provided with L2 data in some form and required to perform some operation on or with it, the purpose of which is to
Rahman Sahragard, Ali Kushki & Ehsan Ansaripour

arrive at an explicit understanding of some linguistic properties of the target language'. Fotos (1993) adds that based on this theory, their consciousness being raised of a particular feature through formal instruction, learners keep remaining aware of intended features and notice it in later exposures. As Ellis (2010) puts it, a consciousness-raising task makes language itself the content by inviting learners to discover how a grammatical feature works for them. In his characterization of a consciousness-raising task, Ellis (2010) counts the four ensuing elements:

1. There is an attempt to *isolate* specific linguistic features for focused attention.
2. The learners are provided with data that illustrate the targeted feature and they may also be provided with an *explicit rule* describing or explaining the feature.
3. The learners are expected to utilize *intellectual effort* to understand the targeted feature.
4. Learners may be optionally required to *verbalize* a rule describing the grammatical feature (Ellis, 1991).

The treatment the participants of the study received had to a great extent the features Ellis (1991) mentioned above. Firstly, the relativizers and their antecedents were isolated for focused attention. Then, the participants were exposed to the data and were expected to utilize intellectual efforts to induce the target feature. In cases where they failed to figure out how the structures worked, they were provided with explicit rules.

One more justification for the significant findings of the study comes from “input flood” or “input enhancement” debate (Larsen-Freeman, 2003; Ellis, 2010; Van Patten & Benati, 2010). Larsen-Freeman (2003) puts it that one of the less explicit means of focusing students’ attention on form is to “underscore, use boldface, use different forms, use color, and so forth to heighten the saliency of some particular grammatical feature in written texts, thereby presumably drawing learners’ attention” (p.92). Ellis (2003) states that input enhancement consists of tasks that the target feature is 1) frequent and/or 2) salient in the input learners are exposed to. He continues to say that one of the forms this technique can take is written texts in which the intended feature has been graphologically highlighted in one way or another. As it was mentioned earlier, again, the results of the study can be justified because the treatment qualifies as an input enhancement task in that the relativizers and the antecedents were graphologically highlighted. Besides, the target features were both frequent and salient in the input given the nature of the corpora they were treated with. The results of the study are similar to that of White (1998), Jourdenais et al. (1996), and Trahey and White (1993). White (1998) tried to draw students’ attention to grammatical forms such as pronouns by printing them in italic or bold face. He found that this was an
effective technique, but not for all of the students. In another study, Jourdenais et al. (1996) found that if English-speaking learners of 2 Spanish had studied texts with the intended forms graphologically highlighted, they were more likely to refer explicitly to preterit and imperfect verb forms. Additionally, Trahey and White (1993) investigated whether enriched input was sufficient to enable French-speaking of L2 English to learn the positions where adverbs in English can be inserted. The results of their study showed that the learners could learn the SAV (Subject-Adverb-Verb) position but failed to “unlearn” the SVAO one. The results of the previous study were compared with those of White (1991) and it was found that the input flood was much less effective when it came to helping learners notice the ungrammaticality of SVAO. Ellis (2003) concludes tentatively from the study that input enrichment may be helpful in contributing learners acquire totally new L2 features but not very effective in enabling them to dispense with incorrect rules that have entered their interlanguage. The results of the present study prove Ellis wrong in this case because the participants of the study not only learned to use alternative relativizers instead of their two default ones (which & that), but also they succeeded in “unlearning” the misapplication of which and that, as was proved form the increase in the number of which- that insertions and also, self-correction documented. One more point which should be added here and confirms the results of the study is the seminal work by Ellis (2002). Having reviewed the literature on the effects of frequency on different aspects of language, she concludes that language processing is closely tuned to input frequency. She adds that this has profound implications for theories of language acquisition in the sense that language learning is exemplar based. Ellis also maintains that underlying the fluent use of language is a massive repertoire of previously experienced utterances not just abstract rules and structures.

Another justification for the results of the study with nearly the same theme proposed by Ellis (1990) comes from Schmidt (1990, 1994, 2002). He distinguishes input or perceived information from intake or the information noticed by a learner. He asserts that noticing linguistic forms is of paramount importance for their later processing. Specifically, he believes that the only linguistic elements in the input that learners can acquire are those elements that they notice. By noticing, Schmidt means learners’ conscious attention. From this position it would follow that instruction would be beneficial as it would direct learners’ attention to linguistic features (Bot et al., 2005; Ellis et al., 2009; Izumi, 2002; Izumi, 2003; Larsen-Freeman, 2003; Robinson, 1995; Saville-Troike, 2006; Schmidt, 2002; Van Patten & Benati, 2010). Schmidt’s proposal springs from his self analytical study (Schmidt & Frota, 1986) when he was learning Portuguese in Brazil. He found that those linguistic features which appeared in his speech were nearly those to which he had noticed once. Therefore, based on his study, only is formal instruction insufficient for subsequent uses of a feature. Rather, it was only after Schmidt noticed the
intended form in the input that the form eventually showed up in his own production. Schmidt (1990) lists the following features as likely contributors to the degree of noticing or awareness which will occur: 1) Frequency of encounter with items, 2) Perceptual saliency of items, 3) Instructional strategies that can structure learner attention, 4) Individuals’ processing ability (a component of aptitude), 5) Readiness to notice particular items (related to hierarchies of complexity), 6) Task demands, or the nature of activity the learner is engaged in. The data the participants were exposed to meet to some extent the features Schmidt enlists above as factors affecting noticing to occur. First, since the data were in the form of corpora, naturally frequency of encounter was enhanced. In terms of saliency, since the data were frequently exposed, so they were more likely to be in the spotlight and consequently, had more attention. Regarding task demands, it should be mentioned that the participants were demanded to focus on the highlighted features in the text. Based on what was mentioned, it can be justified that since all the circumstances for noticing to happen were met, the participants could learn the intended structures.

5 Conclusion

The study was carried out to see how corpus data can be used to address the learning problems Iranian Learners of English had with English relative clauses due to the mismatch existing between Persian and English. As discussed above, the results of the study showed that corpora can be well introduced into our classrooms to contribute learners overcome the learning problems they have when learning English, in this case relative clause types. The study showed that corpora can be an effective tool to enhance the frequency of the target feature in order to make them notice these features and, consequently, their learning process become more convenient. Surely, learning a new language is a serious challenge and fulfilling it requires great personal effort and commitment. But, this personal mobilization should not be construed as leaving the whole process to learners themselves. Instruction can have a lot to propose, as our study proved. It was a new experience of using corpora in a classroom context and its results were promising. Therefore, it is recommended that teachers at the frontiers of teaching employ the data from corpora to help their grappling students have a less painful experience with language learning. They can use corpora to address different aspects of language like vocabulary, syntax, semantics, and pragmatics. Specifically, our study was focused on a portion of syntax and it is hoped that further studies add to the validity of our study.
The Application of Corpora in Teaching Grammar:
The Case of English Relative Clause

References
