Abstract

Due to their structural complexity, English relative clauses (RCs) are difficult to acquire for EFL learners. This study tested the predictions of seven major hypotheses proposed on the difficulty order of SS, SO, OO, and OS RCs for Persian EFL learners with different levels of English proficiency. Data was collected from 39 university students aged between 18 and 22 who performed a sentence comprehension task which consisted of 20 items involving reversible animate head nouns, with 5 items representing each of the SS, SO, OO, and OS RCs. Results showed that the determining factor in the difficulty order of the RCs for Persian EFL learners is the role of the head noun in the RC rather than the position of relativization, as some hypotheses predict. Moreover, Persian EFL learners opt for a linear parsing strategy in processing RC structures. Besides, proficiency level did not bring about a drastic change in the difficulty order of the RCs.

Keywords: English Relative Clauses, Embeddedness, Focus, Sentence Comprehension Task, Persian learners of English

1. Introduction

Investigating the processing of RC structures has proved to be a useful avenue in psycholinguistics to elucidate processing difficulties (Gibson & Wu, 2008). One reason for this interest in RCs is that they are universal in languages of the world, have unique syntactic properties, and are frequent in everyday use of language (Izumi, 2003). A second reason may lie in the fact that RCs present a major obstacle for both first and second language learners. Furthermore, they include recursion, i.e., embedding one instance of a category inside another instance of that category, which is one of the most
distinctive features of language as a cognitive system (Gibson, Desment, Grodner, Watson, & Ko, 2005, p. 314).

Due to their structural complexity, English relative clauses (RCs) are particularly difficult to process for EFL learners. Various hypotheses have been put forward to predict the difficulty order of different RCs. The present study aims at exploring the difficulties Persian EFL learners face in processing English RCs by putting to test the predictions of Keenan and Comrie’s (1977) Noun Phrase Accessibilities Hierarchy Hypothesis (NPAH), Keenan’s (1975) Relativized Subject Accessibility, Kuno’s (1975) Perceptual Difficulty Hypothesis (PDH), Hamilton’s (1994) SO Hierarchy Hypothesis (SOHH), Sheldon’s (1974) Parallel Function Hypothesis, Perspective Shift Hypothesis proposed by MacWhinney and Pleh (1988), and Slobin’s (1973) Non-interruption Hypothesis.

2. Literature Review

Based on their studies on about fifty languages, Keenan and Comrie (1977) proposed the Noun Phrase Accessibility Hierarchy (NPAH) which, according to Fox (1987), is the most robust typological interpretation of RC constraints to date. NPAH is an implicational scale for the relativizability of different grammatical roles according to which all languages adhere to the following scale based on which RCs formed on the subject are hypothesized to be the easiest, while those on the object of a comparative are the most difficult:

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

SU (Subject) the boy that came
DO (Direct Object) the boy that Jane saw
IO (Indirect Object) the boy that she spoke to
Oblique (Object of Preposition) the boy that she sat near
GEN (Genitive) the boy whose mother came
COMP (Object of Comparative) the boy that she is taller than

Although NPAH has received support from research in both L1 (Gibson, 1998; Gibson & Schutze, 1999) and L2 (Doughty, 1991; Gass, 1979; Izumi, 2003; O’Grady, 1999), the applicability of its predictions to all languages of the world (e.g., pre-nominal RC languages like Japanese and Chinese) remains an open question (Ozeki & Shirai, 2007).
Also, a study of RCs in sixteen ergative languages by Hsiu-chuan (2000) reveals that the NPAH is not able to account for the relativization in ergative languages. Researchers like Tarallo and Myhill (1983) have also asserted that NPAH can’t predict the acquisition of RCs appropriately and that the acquisition of RCs is determined on the basis of the distance between head nouns and their trace, not the function of relative markers.

It should be noted that levels such as "SU" on the hierarchy do not represent their positions in sentences, because "SU", for example, can take two positions depending on the position of the head noun in the matrix clause; the NPAH is concerned only with the functions of relative pronouns within the RC, not the functions of relativized noun phrases in the matrix clause. But since, in languages like English the hypothesized comprehension strategies can not be examined separately owing to the interrelationship between certain structural features (e.g., the case of the head and the position of the relative clause), researchers have studied the effect of both embeddedness (i.e., the position of relativization) and focus (i.e., the role the head noun in the RC) on RC processing. Based on this account, Sheldon (1974) has categorized RCs into the four structures in (1). It should be noted that the first letter stands for the position of the RC in the matrix clause and the second letter stands for the function of the head noun in the RC.

(1)

SS: The boy that hits the kid watches the girl.
SO: The boy that the kid hits watches the girl.
OO: The boy hits the kid that the girl watches.
OS: The boy hits the kid that watches the girl.

Relativized Subject Accessibility or RSA (Keenan, 1975) hypothesizes that relativized subjects are more accessible than relativized objects. Therefore, according to this hypothesis, SS and OS types should be easier than SO and OO. Perceptual Difficulty Hypothesis (PDH) proposed by Kuno (1975) asserts that processing center embedded RCs is perceptually more demanding than right branching RCs and on this account, OS and OO should be easier than SS and SO (Doughty, 1991; Ioup & Kruse, 1977; Schumann, 1980). Center embedding is predicted to pose greater difficulty since it interrupts visual processing, but right or left embedding is thought to aid the short-term memory by the absence of interruption (Izumi, 2003). Evidence from the L2 context
supporting PDH comes, to name a few, from Doughty (1991), Ioup and Kruse (1977) who investigated the grammaticality judgments of Spanish, Persian, Japanese, Arabic, and Chinese learners of English on the four RC types (OS, OO, SO, SS), Prideaux and Baker (1986), Schumann (1980) who examined the acquisition order of RCs in the speech production of five Spanish learners of English, and Wong (1990) who investigated 170 English compositions by EFL students in a Hong Kong secondary school. A criticism leveled at PDH is that it accounts only for the position of the RC, right- or center-embedded, but not for the grammatical function of relative markers, subject or object, thus ignoring the second letter in OS, OO, SO, and SS pairs. More specifically, it fails to compare OS with OO, or SS with SO.

Hamilton’s (1994) SO Hierarchy Hypothesis (SOHH) takes into account both the function of the head noun in the matrix clause and the function of the relative pronoun within the RC and bases the order of difficulty on the notion of processing discontinuity which is created by the interruption of the main clause by the RC, i.e., center-embedding, and by phrasal boundaries within the RC that separate the relative pronoun and the trace created by relativization. Hamilton (1994) supported his hypothesis by data from adult ESL learners using a sentence combination test. The order of difficulty predicted by SOHH is: OS > OO = SS > SO. Thus, SS and OO are of the same level of difficulty since they contain three discontinuities. SO is predicted to be the most difficult because it contains four discontinuities, and OS is predicted to be the easiest. The sentences in (2) illustrate the point.

(2)

SS: The boy [CP that [IP__ [VP sees the girl]]] likes the artist
3 discontinuities (2 within the RC and 1 caused by center-embedding)
SO: The boy [CP that [IP the girl [VP sees __ ]]] likes the artist
4 discontinuities (3 within the RC and 1 caused by center-embedding)
OO: The boy sees the girl [CP that [IP the artist [VP likes __ ]]]
3 discontinuities (3 within the RC)
OS: The boy sees the girl [CP that [IP __ [VP likes the artist]]]
2 discontinuities (2 within the RC)
However, the validity of treating in the same way the discontinuity caused by center embedding and the discontinuities created by phrasal boundaries within the RC is not certain since these two types of discontinuities may not carry the same weight in terms of processing burden (Izumi, 2003). It may be on these grounds that researchers have limited their focus to discontinuities within the RC itself and specifically hypothesized about the processing difficulties in terms of the filler-gap distance within the RC. Specifically, the distinction between subject and object RCs which “has formed the basis for investigations in virtually every area of psycholinguistics” (MacDonald et al., 2009, p. 251), has been accounted for by various distance-based hypotheses like the filler-gap hypothesis proposed by Wanner and Maratsos (1978) and supported by several studies (e.g. de Villiers et al., 1979; Hawkins, 1987), and by O’Grady’s (1999) Structural Distance Hypothesis (SDH) which is based on the number of intervening nodes crossed between the gap and the head, as well as by Gibson’s (1998) Dependency Locality Theory (DLT) which is based on integration and storage costs as the distance between the head and the gap increases.

Parallel Function Hypothesis or PFH (Sheldon, 1974) predicts that co-referential NPs that have the same grammatical function in their respective clauses (SS, OO) are easier to process than co-referential NPs with different grammatical functions (SO, OS). In other words, Sheldon (1974) notes that when the antecedents and relative pronouns have the same function (SS, OO), they are easier than when they do not (SO, OS). Research supporting PFH comes from Doughty (1991), Ioup and Kruse (1977), and Schumann (1980). However, many researchers (e.g., Bowerman, 1979; Gass & Ard, 1980; Houston, 1978; Priadeux & Baker, 1986; Sadighi, 1994; Tavakolian, 1977) as well as Sheldon (1977) herself, comparing adults’ and children’s comprehension of RCs, have failed to provide consistent support for PFH.

Defining perspective as the pragmatic counterpart to the formal category of subject, MacWhinney and Pleh (1988) argue that the unmarked processing tendency of speakers and listeners is to see themselves as actors in the world. Their Perspective Shift Hypothesis (PSH) asserts that when the process of perspective sharing is disrupted by interruptions, communication can break down. Based on this theory, processing resources are required to shift the perspective of a clause, where the perspective of a clause is taken
from the subject of the clause. Processing SS RCs like (3a) requires no perspective shifts, because the matrix subject is also the subject of the RC, so that both clauses come from the same perspective. A SO RC as in (3b) requires two perspective shifts: a shift from the perspective of the matrix subject to the subject of the RC; and another from the perspective of the subject of the RC back to the matrix subject after the RC is processed. Thus, SO is more complex than SS.

(3)

a. SS: The boy [that likes the girl] sees the cat.

b. SO: The boy [that the girl likes] sees the cat.

PSH predicts OO RCs to be more difficult than SS RCs because there is one shift of perspective in an OO RC *The boy likes the girl that the cat sees*. In this sentence, perspective begins with *the boy* but then shifts to *the cat* at the end. Based on PSH, in OS structures only one shift of perspective is required, which occurs within the RC. Thus, the order of difficulty predicted by this hypothesis is: SS > OS = OO > SO which, as MacWhinney (2005) claims, has received support from studies of both acquisition and adult processing. Recent evidence shows that PSH may apply in English RCs but does not apply in processing Chinese RCs (Hsiao & Gibson, 2003).

In a similar vein, Slobin’s (1973) Non-interruption Hypothesis (NIH) which can be applied to a matrix sentence as an anti-interruption constraint and to an RC as preference for canonical word order is based on Closure and Normal Form Strategies which have received empirical support in a series of studies conducted by Prideaux and Baker (1986). Closure (Bever, 1970) posits that, because of memory limitations, language users tend to expect that sentences or phrases will be closed as soon as possible. Accordingly, a non-interrupted clause is easier to comprehend than the one which is interrupted because a non-interrupted clause allows early closure. Hence, sentences which are center-embedded are more difficult than sentences in which RCs are right-branching. Normal Form strategy specifies another expectation held by language users that a given unit will be in its usual, unmarked, canonical form. This NVN strategy (Bever, 1970) predicts that any noun-verb-noun corresponding to subject-verb-object will be easier to process than the same sequence corresponding to other grammatical relations. Based on this account, as
the examples in (4) illustrate, OS and OO RCs are easier than SS and SO RCs since they are not interrupted and can thus be processed by the NVN strategy. Also, SO should be the most complex RC type as it contains a VV sequence with little indication of the roles of the associated nouns. This strategy is compatible with Markedness (Prideaux & Hogen, 1993), on the basis of which a clause in an unmarked word order is easier to process than one which is in a marked word order.

(4)

SS: The boy that hits the kid watches the girl.
   N V N V N

SO: The boy that the kid hits watches the girl.
   N N V V N

OO: The boy hits the kid that the girl watches.
   N V N N V

OS: The boy hits the kid that watches the girl.
   N V N V N

Hui Yin (2006), conducting an RC acceptability judgment task, observed that RCs with non-interruption are really easier to process than RCs with interruption in Mandarin Chinese. Sheldon (1974), however, came up with findings that called for reconsidering Slobin’s (1973) claim that a sentence with interruption would be more difficult to process than a sentence without interruption.

Table 1 presents the predictions of the seven major hypotheses mentioned above.

Table 1

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Prediction*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun Phrase Accessibility Hierarchy</td>
<td>SU &gt; DO &gt; IO &gt; PREP &gt; GEN &gt; COMP</td>
</tr>
<tr>
<td>Keenan and Comrie (1977)</td>
<td>SS &amp; OS &gt; OO &amp; SO</td>
</tr>
<tr>
<td>Relativized Subject Accessibility</td>
<td>SS &amp; OS &gt; OO &amp; SO</td>
</tr>
<tr>
<td>Keenan (1975)</td>
<td></td>
</tr>
<tr>
<td>Perceptual Difficulty Hypothesis</td>
<td>OO &amp; OS &gt; SS &amp; SO</td>
</tr>
<tr>
<td>Kuno (1975)</td>
<td></td>
</tr>
</tbody>
</table>
SO Hierarchy Hypothesis (SOHH)  OS > SS = OO > SO  
Hamilton (1994)  
Parallel Function Hypothesis  SS & OO > SO & OS  
Sheldon (1974)  
Perspective Shift Hypothesis  SS > OS = OO > SO  
MacWhinney & Pleh (1988)  
Non-interruption Hypothesis  OO & OS > SS & SO  
Slobin (1973)  

*The greater than sign (>) implies “easier to comprehend than”.

2.1 Studies on English RCs in L2 Context

Sadighi and Jafarpur (1994) examined the instructional materials of 104 Persian EFL learners to obtain the frequency of the types of restrictive RCs in their input. The frequency list was then correlated with the difficulty order obtained from the subjects' performance on a test of RC comprehension. The results did not reveal any significant relationship. The difficulty order, however, showed close correspondence with NPAH (Keenan & Comrie, 1977).

Takashima et al. (1994) used a translation task in which students were asked to evaluate the difficulty of translating different types of English RCs into Japanese, as well as a questionnaire asking them to detect the subject and/or the object of RCs. The results showed that SO seems to be the most difficult among the four types, followed by OO, SS and OS, confirming the findings of a study by Ina (1982), and the claims of Hamilton (1994), MacWhinney and Pleh (1988), Keenan and Comrie (1977), Sheldon (1974), Kuno (1975), Keenan (1975), and Slobin (1973).

Takashima (2000; cited in Takashima et al., 1994) investigating the correspondence between the order of difficulty that Japanese high school and university students perceive with respect to English RCs and the frequency count order from a corpus of written data, came up with OS > SS > OO > SO.

Romaine (1984) collected speech data from 24 primary school children and, based on the overall frequencies of RCs produced by the children, reported the following hierarchy: OS > OO > SS > SO. He concluded that the factor of embeddedness is clearly
the one which carries the most weight, with object RCs greatly preferred over subject ones, lending support to Non-interruption Hypothesis (Slobin, 1973) and Perceptual Difficulty Hypothesis (Kuno, 1975).

In another study in Chinese, (Hsin, 2005) data collected from 84 Chinese high school seniors performing two different tasks, a sentence comprehension task and a combination task, were analyzed and full support was found for NPAH and PDH and partial support for SOHH. Subjects with higher English proficiency had similar performance to those with lower proficiency on the sentence comprehension test but a significantly better performance on the combination test.

3. Methodology

This study seeks to explore the difficulty Persian EFL learners encounter in the comprehension of English SS, SO, OS, and OO RCs, thus concentrating on both the effect of focus, i.e., the function of the head noun in the RC, and that of embeddedness, i.e., the position of the RC in the matrix sentence, on sentence processing. More specifically, the study aims to test the predictions of the above-mentioned hypotheses. Attempt is made to present a few principles to explain the difficulty order of SS, SO, OO, and OS RCs for EFL learners with different levels of English proficiency. Finally, the extent of L1 transfer is tackled to see how the native language of EFL learners influences their comprehension of English RCs.

3.1. Participants

Thirty nine male and female Persian native speakers aged between 18 and 22 majoring in English Translation in two universities were recruited for the present study (6 subjects were excluded due to reasons mentioned below). The participants were divided into two groups based on their level of English proficiency. The high proficiency group consisted of 19 subjects, and the low proficiency group consisted of 14 subjects. They received extra course credit for their participation in a single testing session.

3.2. Materials

The participants in the study performed a sentence comprehension task which
consisted of 20 items, with 5 items representing each of the four RC types (SS, SO, OS, and OO). All the verbs in both the RC and the matrix clause were transitive and used in the present tense. All the noun phrases were animate to control for possible animacy effects as animacy of the subject or object of a clause has been shown to affect comprehension to a great extent (e.g. Gibson, et al., 2005). Animacy of all the noun phrases and, therefore, their reversibility, prevented the participants from using the semantic cue without using their grammatical knowledge. Also, all the noun phrases had the same person and number to factor out possible cues from verb agreement. Nelson’s Quickcheck Test (50 items) was used to assess the participants’ English proficiency in order to examine whether learners with different proficiency levels would perform differently on RC comprehension task.

3.3. Procedure

Participants first took Nelson’s Quickcheck Test, then performed the sentence comprehension task. Prior to the task, a few sample training items were presented to make sure the participants knew the procedure. At this stage, there was sufficient interaction between the experimenter and the participants to ensure they knew how to proceed. However, during the actual test phase no feedback was given. The entire procedure lasted about 45 minutes, 30 minutes for Nelson’s Quickcheck Test, and 15 minutes for the sentence comprehension task. To complete the comprehension task, the participants were required to read each sentence and identify the subject and object of the verb in the matrix clause and those of the verb in the RC. A sample test item is presented in (5) below. To complete this item, the subjects were supposed to provide the subject and object of the verb watch and those of the verb push and write them in blank spaces.

(5) The dog that watches the cat pushes the horse.

(a) [------------ watches---------]   (b) [-------------pushes----------]

The blanks are referred to as positions. In this SS RC example, the first position is referred to as SSES, the first two letters refer to the RC type and the second two letters refer to the Embedded Subject position. The second position is referred to as SSEO, EO referring to the Embedded Object position, the third position is referred to as SSMS, MS referring to the Matrix Subject position and the last position belongs to SSMO, MO
referring to the Matrix Object position. As for the scoring system, a score of 1 was given for each correct answer and a score of 0 for each incorrect answer.

4. Data Analysis

The participants’ scores on Nelson’s Quickcheck Test were pruned, and the outliers were identified and excluded (42, 42, 41 as the highest and 7, 9 and 10 as the lowest scores). In this way, the analyzable data came from only 33 participants. The set of scores was then tested for normality. Both the Kolmogorov-Smirnov and the Shapiro-Wilk tests showed that the set had a normal distribution. The mean for Nelson’s Quickcheck Test was 27.8 and the standard deviation was 7.45. The scores were divided into two groups of high and low; those who scored above the mean were considered as high in proficiency and those who scored below the mean were considered as low in proficiency. To make sure that the two groups were different from each other, an independent samples t-test was conducted. The results showed that the high proficiency group with a mean of 33.16 was significantly better than the low proficiency group with a mean of 20.57 (t$_{31}$ = 8.88, p = 0.000).

5. Results and Findings

As Table 2 below shows, regardless of the proficiency level, performance on the SO type is the worst (mean = 12.58) and on OS the best (mean = 18.48).

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>SO</th>
<th>OS</th>
<th>OO</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>17.37</td>
<td>15.26</td>
<td>18.16</td>
<td>16.53</td>
</tr>
<tr>
<td>SD</td>
<td>2.65</td>
<td>4.90</td>
<td>2.03</td>
<td>4.15</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>16.00</td>
<td>8.93</td>
<td>18.93</td>
<td>12.29</td>
</tr>
<tr>
<td>SD</td>
<td>1.617</td>
<td>4.62</td>
<td>1.64</td>
<td>4.12</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>16.79</td>
<td>12.58</td>
<td>18.48</td>
<td>14.73</td>
</tr>
<tr>
<td>SD</td>
<td>2.34</td>
<td>5.68</td>
<td>1.89</td>
<td>4.60</td>
</tr>
</tbody>
</table>
A repeated measures ANOVA was conducted, with sentence type (SS, SO, OS, OO) as the within-group variable, and proficiency level (high, low) as the between group variable. The results showed that sentence type has a significant effect ($F_{(3, 29)} = 34.2$, $p = .000$). Proficiency level, too, showed a significant effect ($F_{(1, 31)} = 9.45$, $p = .000$); the interaction effect was significant as well ($F_{(3, 93)} = 10.72$, $p = .000$). Pairwise comparisons showed that a significant difference exists among all four RC types. Thus, the analysis shows that the hierarchy of difficulty of the four RC types is as follows: OS (Mean = 18.48) > SS (Mean = 16.79) > OO (Mean = 14.73) > SO (Mean = 12.58). This hierarchy holds true across proficiency levels as well. For the high proficiency group, the hierarchy is: OS (Mean = 18.93) > SS (Mean = 16) > OO (Mean = 12.29) > SO (Mean = 8.93). And for the low proficiency group, the hierarchy is: OS (Mean = 18.16) > SS (Mean = 17.37) > OO (Mean = 16.53) > SO (Mean = 15.26).

A question which was raised here was whether the effect of embeddedness was greater than that of function or not. The mean of the participants’ performance on subject embedded RCs (SS (Mean = 16.79), SO (Mean = 12.58)) is 14.68 and the mean of their performance on object embedded RCs (OS (Mean = 18.84), OO (Mean = 14.73)) is 16.60. The difference between the two means is 1.92. Also, the mean of the participants’ performance on subject focus RCs (SS (Mean = 16.79), OS (Mean = 18.48)) is 17.63 and the mean of their performance on object focus RCs (SO (Mean = 12.58), OO (Mean = 14.73)) is 13.65. Thus, the difference between the two means is 3.98. A comparison of the two mean differences (1.92 & 3.98) shows that focus carries a greater weight than embeddedness.

To analyze the data on the participants’ ability to identify correctly the subject and object of the verb in the matrix clause and those of the verb in the RC, the four positions were designated as Matrix Subject, Matrix Object, Embedded Subject, and Embedded Object. Different repeated measures ANOVAs were conducted for each RC type to see which position is the most problematic one for L2 learners in each RC sentence type.

The following table shows descriptive statistics for each of the four positions in the SS RC type.
Table 3

*Descriptive statistics for the four positions in SS RCs*

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>SSES</th>
<th>SSEO</th>
<th>SSMS</th>
<th>SSMO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Mean</td>
<td>5</td>
<td>4.78</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.00</td>
<td>0.80</td>
<td>1.49</td>
</tr>
<tr>
<td>High</td>
<td>Mean</td>
<td>4.42</td>
<td>4.36</td>
<td>3.73</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.34</td>
<td>1.34</td>
<td>1.32</td>
</tr>
<tr>
<td>Total</td>
<td>Mean</td>
<td>4.66</td>
<td>4.54</td>
<td>2.72</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.05</td>
<td>1.14</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Example SS RC sentence: “The boy [that sees the girl] likes the artist”

[... SSES ... sees ... SSEO ...]          [... SSMS ... likes ... SSMO ...]

As can be seen from the means, the best performance is when subjects were supposed to identify the matrix object (Mean = 4.84) and the worst one is when they were supposed to identify the matrix subject (Mean = 2.72). To see if the differences are significant or not, a repeated measures ANOVA was conducted. Results showed that position had a main effect on the subjects’ performance ($F(3, 29) = 36.08, p = .000$), but proficiency level did not ($F(1, 31) = 2.92, p = .09$). As for the interaction between the two variables, the effect turned out to be significant ($F(3, 93) = 14.48, p = .000$). Pairwise comparisons showed that Matrix Subject position was the most difficult one. The difficulty order of the four positions in SS RC sentences is: SSMO (Mean = 4.84) = SSES (Mean = 4.66) = SSEO (Mean = 4.54) > SSMS (Mean = 2.72). This hierarchy holds true for the low proficiency group as the order is: SSMO (Mean = 4.85) = SSES (Mean = 5) = SSEO (Mean = 4.78) > SSMS (Mean = 1.35). But for the high proficiency group, the order is: SSMO (Mean = 4.82) > SSMS (Mean = 3.73) = SSES (Mean = 4.42) = SSEO (Mean = 4.36). Thus, for this group, SSMS is only significantly more difficult than SSMO, not the other two positions, i.e., SSES and SSEO.

The following table shows descriptive statistics for each of the four positions in the SO RC type. The difficulty order of the four positions in SO RC sentences is: SOMO (Mean = 4.75) > SOMS (Mean = 2.90) = SOES (Mean = 2.57) = SOEO (Mean = 2.33).
Table 4

Descriptive statistics for the four positions in SO RCs

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>SOES</th>
<th>SOEO</th>
<th>SOMS</th>
<th>SOMO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Mean</td>
<td>1.71</td>
<td>1.14</td>
<td>1.50</td>
<td>4.57</td>
</tr>
<tr>
<td>Low SD</td>
<td>1.72</td>
<td>1.70</td>
<td>1.78</td>
<td>0.75</td>
</tr>
<tr>
<td>High Mean</td>
<td>3.21</td>
<td>3.21</td>
<td>3.94</td>
<td>4.89</td>
</tr>
<tr>
<td>High SD</td>
<td>1.84</td>
<td>1.84</td>
<td>1.43</td>
<td>0.31</td>
</tr>
<tr>
<td>Total Mean</td>
<td>2.57</td>
<td>2.33</td>
<td>2.90</td>
<td>4.75</td>
</tr>
<tr>
<td>Total SD</td>
<td>1.92</td>
<td>2.04</td>
<td>1.99</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Example SO RC sentence: “The boy [that the girl sees] likes the artist”

[… SOES … sees… SOEO …]          [… SOMS … likes … SOMO …]

As the table shows, the easiest position for the subjects to fill was the Matrix Object (Mean = 4.76). To examine the significance of the differences, a repeated measures ANOVA was conducted. The results showed that position had a significant effect on the replies (F(3, 29) = 36.95, p = .000); proficiency level, too, showed a main effect (F(1, 31) = 14.15, p = .001); the interaction between the two variables, however, was not significant (F(3, 93) = 5.89, p = .098). Pairwise comparisons showed that the fourth position, i.e., the Matrix Object position, was the easiest one and the second position, i.e., the Embedded Object position, was the most difficult one. For the low proficiency group, the difficulty order is: SOMO (Mean = 4.57) > SOES (Mean = 1.71) = SOMS (Mean = 1.50) = SOEO (Mean = 1.14). For the high proficiency group, the order is: SOMO (Mean = 4.89) > SOMS (Mean = 3.95) > SOES (Mean = 3.21) = SOEO (Mean = 3.21). Thus, for both groups, SOMO is the easiest position and SOEO is the most difficult one.

As for OS RC type sentences, as Table 5 shows, the Matrix Subject position was the easiest one for the participants (Mean = 4.94) and the Embedded Subject position was the most difficult one (Mean = 3.97).
Table 5

Descriptive statistics for the four positions in OS RCs

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>OSMS</th>
<th>OSMO</th>
<th>OSES</th>
<th>OSEO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Mean</td>
<td>5</td>
<td>4.85</td>
<td>4.14</td>
<td>4.92</td>
</tr>
<tr>
<td>SD</td>
<td>0.00</td>
<td>0.53</td>
<td>1.02</td>
<td>0.26</td>
</tr>
<tr>
<td>High Mean</td>
<td>4.89</td>
<td>4.89</td>
<td>3.84</td>
<td>4.52</td>
</tr>
<tr>
<td>SD</td>
<td>0.31</td>
<td>0.31</td>
<td>1.06</td>
<td>0.96</td>
</tr>
<tr>
<td>Total Mean</td>
<td>4.94</td>
<td>4.87</td>
<td>3.97</td>
<td>4.69</td>
</tr>
<tr>
<td>SD</td>
<td>0.24</td>
<td>0.41</td>
<td>1.04</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Example OS RC sentence: “The boy sees the girl that likes the artist”
[...OSMS ...sees... OSMO ...]           [...OSES ...likes... OSEO ...]

The repeated measures ANOVA results showed a main effect for position ($F(3, 29) = 18.12, p = .000$); but no effect was found for proficiency level ($F(1, 31) = 1.36, p = .253$); the interaction was not significant either ($F(3, 93) = .913, p = .438$). The results of pairwise comparisons showed that the third position, i.e., the Embedded Subject position, was the most difficult and significantly different from the other three positions. The difficulty order of the four positions in OS RCs is: OSMS (Mean = 4.94) = OSMO (Mean = 4.87) = OSEO (Mean = 4.69) > OSES (Mean = 3.97). The same hierarchy applies to both proficiency groups.

In this part, we go to OO RC type sentences. Table 6 shows descriptive statistics for each of the four positions in the OO RC type.

Table 6

Descriptive statistics for the four positions in OO RCs

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>OOMS</th>
<th>OOMO</th>
<th>OOES</th>
<th>OOEO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Mean</td>
<td>5</td>
<td>4.71</td>
<td>1.42</td>
<td>1.14</td>
</tr>
<tr>
<td>SD</td>
<td>0.00</td>
<td>0.82</td>
<td>2.10</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>4.94</td>
<td>4.94</td>
<td>3.31</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>High</td>
<td>SD</td>
<td>0.22</td>
<td>0.22</td>
<td>1.97</td>
</tr>
<tr>
<td>Total</td>
<td>Mean</td>
<td>4.96</td>
<td>4.84</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.17</td>
<td>0.56</td>
<td>2.20</td>
</tr>
</tbody>
</table>

Example OO RC sentence: “The boy sees the girl [that the artist likes].”

[…OOMS …sees… OOMO …] ![image](image.png) […]OOES …likes… OOEO …]

The repeated measures ANOVA results showed a main effect for position (F(3, 29) = 55.98, p = .000); proficiency level showed a significant effect (F(1, 31) = 8.45, p = .007); the interaction was significant too (F(3, 93) = 7.93, p = .000). The results of pairwise comparisons showed that the third and fourth positions, i.e., Embedded Subject and Embedded Object, were the most difficult ones and significantly different from the other two positions. The difficulty order of the four positions in OO RCs is: OOMS (Mean = 4.96) = OOMO (Mean = 4.84) > OOES (Mean = 2.51) > OOEO (Mean = 2.39). There is no difference in the participants’ performance on the first two positions, but both were observed to be easier than the third and fourth positions. Also, performance on the third position is significantly better than that on the fourth position. As for the low proficiency group, the hierarchy is OOMS (Mean = 5) = OOMO (Mean = 4.71) > OOES (Mean = 1.42) = OOEO (Mean = 1.14), which is exactly the same as the hierarchy for the high proficiency group (OOMS (Mean = 4.94) = OOMO (Mean = 4.94) > OOES (Mean = 3.31) = OOEO (Mean = 3.31). In both cases, the first two positions are significantly easier than the last two positions.

To identify the most difficult positions from among all the 16 positions, a repeated measures ANOVA was conducted for the low proficiency group and the results (F(15, 13) = 37.12, p = 000) showed that performance on positions SSES, OSMS, OOMS, OSEO, SSMO, OSMO, SSEO, OOMO was not significantly different from each other but significantly better than that on SOMO, OSES, and the performance on these two positions was better than performance on SOES, SOMS, OOES, SSMS, SOEO, OOEO. Thus, for the low proficiency group, the easiest positions to fill were SSES, OSMS, OOMS, OSEO, SSMO, OSMO, SSEO, and OOMO, the middle positions were SOMO and OSES, and the most difficult ones were SOES, SOMS, OOES, SSMS, SOEO, and
Table 7

*Descriptive statistics for the 16 positions for the low proficiency group*

<table>
<thead>
<tr>
<th>Position</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSES</td>
<td>5.00</td>
<td>.00</td>
</tr>
<tr>
<td>OSMS</td>
<td>5.00</td>
<td>.00</td>
</tr>
<tr>
<td>OOMS</td>
<td>5.00</td>
<td>.00</td>
</tr>
<tr>
<td>OSEO</td>
<td>4.93</td>
<td>.27</td>
</tr>
<tr>
<td>SSMO</td>
<td>4.85</td>
<td>.36</td>
</tr>
<tr>
<td>OSMO</td>
<td>4.86</td>
<td>.53</td>
</tr>
<tr>
<td>SSEO</td>
<td>4.78</td>
<td>.80</td>
</tr>
<tr>
<td>OOMO</td>
<td>4.71</td>
<td>.83</td>
</tr>
<tr>
<td>SOMO</td>
<td>4.57</td>
<td>.76</td>
</tr>
<tr>
<td>OSES</td>
<td>4.14</td>
<td>1.03</td>
</tr>
<tr>
<td>SOES</td>
<td>1.71</td>
<td>1.73</td>
</tr>
<tr>
<td>SOMS</td>
<td>1.50</td>
<td>1.79</td>
</tr>
<tr>
<td>OOES</td>
<td>1.43</td>
<td>2.10</td>
</tr>
<tr>
<td>SSMS</td>
<td>1.36</td>
<td>1.50</td>
</tr>
<tr>
<td>SOEO</td>
<td>1.14</td>
<td>1.70</td>
</tr>
<tr>
<td>OOEO</td>
<td>1.14</td>
<td>1.88</td>
</tr>
</tbody>
</table>

As for the high proficiency group, repeated measures ANOVA results \( F(15, 18) = 7.70, \ p = 0.00 \) showed that the easiest positions were OOMS, OOMO, SOMO, OSMS, OSMO, and SSMO, the middle ones in the hierarchy were OSEO, SSES, SSEO, SOMS, OSES, and SSMS, and the most difficult ones were OOES, OOEO, SOES, and SOEO. Table shows the results.
Table 8

Descriptive statistics for the 16 positions for the high proficiency group

<table>
<thead>
<tr>
<th>Position</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>OOMO</td>
<td>4.95</td>
<td>.22</td>
</tr>
<tr>
<td>OOMS</td>
<td>4.95</td>
<td>.23</td>
</tr>
<tr>
<td>OSMO</td>
<td>4.89</td>
<td>.33</td>
</tr>
<tr>
<td>OSMS</td>
<td>4.89</td>
<td>.32</td>
</tr>
<tr>
<td>SOMO</td>
<td>4.89</td>
<td>.32</td>
</tr>
<tr>
<td>SSMO</td>
<td>4.84</td>
<td>.37</td>
</tr>
<tr>
<td>OSEO</td>
<td>4.53</td>
<td>.96</td>
</tr>
<tr>
<td>SSES</td>
<td>4.42</td>
<td>1.35</td>
</tr>
<tr>
<td>SSEO</td>
<td>4.36</td>
<td>1.34</td>
</tr>
<tr>
<td>SOMS</td>
<td>3.95</td>
<td>1.43</td>
</tr>
<tr>
<td>OSES</td>
<td>3.84</td>
<td>1.07</td>
</tr>
<tr>
<td>SSMS</td>
<td>3.74</td>
<td>1.33</td>
</tr>
<tr>
<td>OOEO</td>
<td>3.32</td>
<td>1.97</td>
</tr>
<tr>
<td>OOES</td>
<td>3.32</td>
<td>1.97</td>
</tr>
<tr>
<td>SOEO</td>
<td>3.21</td>
<td>1.84</td>
</tr>
<tr>
<td>SOES</td>
<td>3.21</td>
<td>1.84</td>
</tr>
</tbody>
</table>

6. Discussion and Conclusion

As the results of the study show, the hierarchy of difficulty of the four RC types is OS > SS > OO > SO, with the SO type being the most difficult and the OS type being the easiest. The findings of the study can best be accounted for by Hamilton’s (1994) SOHH which is based on the notion of processing discontinuity created by interruptions within both the matrix clause and the RC. The only disparity between SOHH’s prediction and the observations in the present study is identical complexities predicted for SS and OO, which has not been supported by the results of this study. An analysis of Persian RCs based on SOHH shows that the native language of the EFL learners in this study has influenced their comprehension of English RCs. As the sentences in (6) below show, in addition to basic differences between Persian and English RCs noted by Karimi (2001),
there is a sharp difference between them in that branching does not exist in Persian. Persian is a null-subject verb-final language with SOV word order, and therefore, the final constituent in any Persian sentence is a verb phrase. Consequently, RC structures are always embedded within the matrix clause and right-branching is not allowed. Based on a count of discontinuities created by center-embedding, and phrasal boundaries within the RC that separate the relative pronoun and the trace (Hamilton, 1994), Persian SO and OO RCs are predicted to be the difficult ones and SS and OS RCs, the easier ones. Thus, the hierarchy of difficulty of the four RC types in Persian is SS = OS > SO = OO, which is compatible with the hierarchy obtained in the present study (OS > SS > OO > SO) in that SO and OO RCs in both hierarchies are the two more difficult RC types.

(6)

SS: sæg-i CP[ke IP[ VP[gorbe ra mibinæd]] æsb ra hol midæhæd
  dog-RM that cat OBJECT MARKER sees horse OBJECT MARKER pushes
  3 discontinuities (2 within the RC and 1 caused by center-embedding)

SO: sæg-i CP[ke VP[gorbe u ra mibinæd]] æsb ra hol midæhæd
  dog-RM that cat RP OBJECT MARKER sees horse OBJECT MARKER pushes
  4 discontinuities (3 within the RC and 1 caused by center-embedding)

OS: sæg gorbe-i CP[ke IP[ æsb ra mibinæd]] ra hol midæhæd
  dog cat-RM that horse OBJECT MARKER sees OBJECT MARKER pushes
  3 discontinuities (2 within the RC and 1 caused by center-embedding)

OO: sæg gorbe-i CP[ke VP[æsb u ra mibinæd]] ra hol midæhæd
  dog cat-RM that horse RP OBJECT MARKER sees OBJECT MARKER pushes
  4 discontinuities (3 within the RC and 1 caused by center-embedding)

The most obvious finding of the study is that subject focus RCs (OS, SS) are easier than object focus RCs (OO, SO), a finding in line with the predictions of the difficulty order by both Keenan and Comrie’s (1977) Noun Phrase Accessibility Hierarchy and Keenan’s (1975) Relativized Subject Accessibility. Thus, the claim that relativized subjects are more accessible than relativized objects is supported by the findings. However, both Noun Phrase Accessibility Hierarchy and Relativized Subject Accessibility are concerned only with the functions of relative pronouns within the RC,
not the functions of relativized noun phrases in the matrix clause and, therefore, do not make any comparisons between SS and OS or between OO and SO. Nonetheless, as the results of the study show, focusing only on the functions of relative pronouns within the RC can not yield a complete account of the difficulty order of the four RC types since RC structures in which the relative pronouns were of the same functions (SS (Mean = 16.79) vs. OS (Mean = 18.84) & SO (Mean = 12.58) vs. OO (Mean = 14.73)) were not observed to create identical complexities for EFL learners.

A second finding of the study is that center embedding RCs (SS, SO) are not more complex than right branching RCs (OO, OS), which is in contrast with Kuno’s Perceptual Difficulty Hypothesis (1975) and Slobin’s (1973) Non-interruption. It seems that embeddedness alone can not predict the complexity level of RC structures. Furthermore, the results of the present study show that focus carries a greater weight than embeddedness in processing RCs, contrary to Perceptual Difficulty Hypothesis (Kuno, 1975), Non-interruption Hypothesis (Slobin, 1973) and Romaine’s (1984) finding that embeddedness carries the most weight. Also, co-referential NPs that have the same grammatical function in their respective clauses (SS, OO) were not found to be easier to process than co-referential NPs with different grammatical functions (SO, OS). Thus, a further outcome of the study is the rejection of a role for the co-functionality of co-referential NPs in RC structures claimed by Parallel Function Hypothesis (Sheldon, 1974).

The finding that SO is the most difficult RC is fully in line with the predictions of SO Hierarchy Hypothesis (Hamilton, 1994) and Perspective Shift Hypothesis (MacWhinney & Pleh, 1988), and partially supports the prediction of Noun Phrase Accessibility Hierarchy (Keenan & Comrie, 1977), Parallel Function Hypothesis (Sheldon, 1974), Perceptual Difficulty Hypothesis (Kuno, 1975), Relativized Subject Accessibility (Keenan, 1975), and Non-interruption Hypothesis (Slobin, 1973) which predict SO to be one of the two more difficult RCs. The same level of difficulty predicted by Perspective Shift Hypothesis for OO and OS RCs, however, is not supported by the findings since the observation was that OS RCs are more problematic than OO RCs for EFL learners.

The analysis of the subjects’ ability to correctly identify Matrix Subject, Matrix Object, Embedded Subject, and Embedded Object shows that the most difficult positions
from among the 16 positions in the four RC types for both proficiency groups were observed to be OOEO, OOES, SOEO, and SOES, and the easiest ones were OOMS, OOMO, OSMO, OSMS, OSEO, and SSMO.

SO RCs were found to be the most demanding ones for Persian EFL learners due to greater difficulties in identifying SOEO and SOES positions in this RC type. The difficulty order of the four positions in SO RCs was found to be: SOMO (Mean = 4.75) > SOMS (Mean = 2.90) > SOES (Mean = 2.57) > SOEO (Mean = 2.33). As can be seen, the Embedded Object was the most difficult position to fill, followed by the Embedded Subject. This is in line with various distance-based hypotheses (e.g., the filler-gap hypothesis proposed by Gibson’s Dependency Locality Theory, 1998; O’Grady’s Structural Distance Hypothesis, 1999; and Wanner and Maratsos, 1978) and a large body of research which indicates that, in English, object gap RCs as in *The boy that the girl sees* are difficult to understand compared with subject gap RCs (de Villiers et al., 1979; Hawkins, 1987; King and Just 1991; Traxler et al. 2002). It seems that in a SO sentence like *The boy [that the girl sees] likes the artist, the boy* has been taken to be the Matrix Subject and *the girl*, the Matrix Object. The same problem was observed in OO RCs, making it hard for the subjects in this study to identify the Embedded Subject and the Embedded Object. The question which is raised here is why Persian speaking EFL learners face great difficulties in comprehending object RCs. One reason for this difficulty may lie in the fact that Persian allows a personal or a clitic pronoun within the RC, representing the missing head noun. In other words, personal pronouns can be used resumptively in Persian where a gap (example 7 below) might be expected. Example (8) represents a Persian RC in which the pronoun *u*, ‘he’, is used resumptively and example (9) shows a clitic pronoun, *eš* “him”, used resumptively.

(7)  *mærd-i [ke ___ molaqt kærdid] aqay-e Bayat bud*
    ‘The man whom you met was Mr. Bayat.’

(8)  *mærd-i [ke *u* ra molaqt kærdid] aqay-e Bayat bud*
    ‘*The man whom you met him was Mr. Bayat.

(9)  *mærd-i [ke molaqateš kærdid] aqay-e Bayat bud*
    ‘*The man whom you met him was Mr. Bayat.’
Since in English there is no resumptive pronoun in the RC to aid comprehension for Persian EFL learners, they face greater difficulty in understanding SO and OO RCs. This difficulty can also be explained by the possibility that, to comprehend SO and OO RCs, the learners may opt for a linear parsing and, due to their previous experience with the canonical word order in English (i.e., SVO), take the first noun phrase as the subject and once they encounter the second noun phrase, they realize they have to do a relative clause reading rather than a main clause reading. Therefore, it can be argued that the difficulty is due to a reanalysis effect, an account which is compatible with Bever’s (1970) NVN strategy which asserts that a VV or NN sequence makes comprehension more demanding, as well as Gibson’s (1998) Dependency Locality Theory, and Van der Lely’s (1996) Representational Deficit for Dependent Relationship theory which defines the difficulty in terms of inability to assign thematic roles to noun phrases that have moved from their original position and which are interpreted according to their position within the sentence; if this NP is the first NP, it is correctly interpreted as the agent, as is the case for subject RCs in which the order is canonical. However, when it is a theme, an inappropriate agent role is assigned.

The third difficult position in SO RCs which was challenging only for the low proficiency group was the Matrix Subject position (Mean = 2.90) which seems to be complex due to embedding. In *The boy [that the girl sees] likes the artist*, the participants’ inability to identify the embeddness of an RC within the matrix clause has misled them to incorrectly take *the boy* to be the subject of the verb *sees* and, therefore, to ignore the association between *the boy* and the verb *likes* which were separated by embedding. This observation further supports the linearity of the processing employed by the EFL learners. The same phenomenon was observed in SS sentences like *The boy [that sees the girl] likes the artist*, in which the Embedded Object *the girl* seems to have been taken to be both the object of the verb *sees*, which is the correct interpretation, and the subject of the verb *likes*, which is incorrect, misleading the participants in the study to identify the association between *the boy* and the verb *likes* which were separated by embedding.

The most problematic position in OO RCs was OOEO (Mean = 2.39), followed by OOES (Mean = 2.51) which, as mentioned above, shows that the presence of a personal or
a clitic pronoun in Persian object RCs and the absence of such pronouns in English object RCs make it harder for Persian EFL learners to correctly identify the Embedded Subject and Embedded Object in an English OO sentence like The boy sees the girl [that the artist likes].

OS sentences were the least complex English RC type for the participants. They faced little difficulty in correctly identifying the subject and object of, for example, the verb sees in The boy sees the girl [that likes the artist], maybe due to the fact that the matrix clause is a simple canonical English sentence. Since one of the main difficulties the subjects in this study faced seems to be in identifying recursion within the noun phrase and in processing RCs as noun modifiers, it is plausible to argue that, having ignored the relative marker that, they seem to have processed this RC structure as consisting of two conjoined simple sentences, interpreting it as The boy sees the girl, and the girl likes the artist, a finding which further supports the linear nature of RC processing.

Thus, based on the findings of the present study, the principles governing the processing of RC structures by Persian EFL learners can be summarized as follows: 1. Subject focus RCs are easier than object focus RCs. 2. Interruptions within both the matrix clause and the RC should be accounted for in explaining the difficulty order of RCs. 2. Neither focus nor embeddedness alone can yield a complete account of the complexity level of RC structures. 3. Focus carries a greater weight than embeddedness in processing RCs for EFL learners. 4. Center embedding RCs are not more complex than right branching RCs. 5. Parallelism of grammatical functions in RC structures does not count in RC processing. 6. Persian EFL learners opt for a linear parsing strategy in processing RC structures. 7. Word order canonicity makes a difference in processing RCs. 8. Distance-based hypotheses can account for the difficulty order of RC processing. 9. Certain features in EFL learners’ L1 seem to aid or impede their comprehension of English RCs.

References


tedlab.mit.edu/tedlab_website/.../Gibson%20&Wu%20subm.pdf


---

**Hamideh Marefat** is an Associate Professor at the University of Tehran, Faculty of Foreign Languages and Literatures. Her research interests include second language acquisition and psycholinguistics.

**Ramin Rahmany** is a PhD student in TEFL at the University of Tehran. His research interests include language learning and sentence processing.