The Processing of Topicalization in Persian and English

Leila Azizi
Islamic Azad University, Takestan Branch
Shahla Roshani Ghale Sheikhi
Payam Noor University

Abstract
The present study sought to investigate whether there is any difference between topicalization processes in English and Persian structures among EFL learners. 50 female students of Jahad Daneshgahi English Language Institute in Qazvin participated in this study. All of the participants were native speakers of Persian studying in Jahad Daneshgahi English Language Institute in Qazvin, Iran. The participants were administered three tests. A multiple choice test for English language proficiency test (PET) to determine the level of the participants, a multiple choice English Topicalization test and Persian Topicalization test to measure the possible effect of topicalized structures on subjects' comprehension and to measure the participants' knowledge of topicalized structures. The same tests were presented to participants of advanced and intermediate level. The Wilcoxon Signed Rank Test was used to answer the research question of this study. The findings showed that the advanced group performed better on the English Topicalization test. There was a significant dependency between the English language proficiency level and comprehension of Topicalized structures. Then, the result of Persian test showed that both groups cannot process Persian topicalized sentences well. Accordingly, the null hypothesis that states that there is no significant difference between topicalization process in English and Persian is rejected and we can claim that there is significant difference between topicalization process in English and Persian. The findings of the present study may have implications for L2 learners and teachers.

Key Terms: Topicalization, Processing

2. Introduction
Why do speakers make the syntactic choices that they do? Answering this question requires us to understand both the speaker’s array of options and the manner in which these options present themselves. This understanding relies in turn upon our ability to delimit the conditions—both necessary and sufficient—which constrain the use of each option. Where a language offers
different means of syntactic expression for a given predicate-argument structure, it seems natural to represent this state of affairs by a rule which mediates between the alternates, and yet this analytic mode has been pursued vigorously only in the domain of verbal ‘linking rules’, whose productivity is typically so highly constrained by verb semantics that the ‘rules’ are most appropriately viewed as generalizations over semantic classes in the lexicon (Pinker, 1989. Levin, 1993). When there is no basis for proposing a derivational relationship between two or more syntactic patterns or any reasonable means of cross indexing the patterns (e.g., in a lexical entry), there is no obvious way to model the relationship between two options afforded by the grammar. Accordingly, functional syntacticians have tended to focus either on the use conditions associated with particular pragmatically motivated constructions (Michaelis & Lambrecht, 1996) or on pragmatic constraints attributable to classes of sentence types, e.g., proposing constructions (Ward, 1999. Birner & Ward, 1998).

Topicalization processes are regarded by Fillmore as 'devices for isolating one constituent of a sentence as "topic" (Fillmore, 1968, p. 57). It is easy to extend this notion to complex lexical items, especially if these are considered as derived from underlying sentences, as is done by Marchand. With simple lexical items, it seems impossible, at first glance, to single out constituents. However, the distinction between morphologically complex and simple items can be seen as fairly superficial. The idea that apparently simple linguistic elements can be broken down into further components is not new. It can be traced back to the distinctive feature analysis of the Prague school of phonology which considered the phoneme as a bundle of simultaneously present features. Topicalization will here be used to denote a process of both foregrounding of information and selection of the 'topic of information, which is a process which singles out certain elements in a sentence and makes them the 'topic on which some 'comment is made. The
separation of elements permits more specific modification, as in the case of with almost, or in the cleft sentences It wasn't John who was killed (by Harry), It was John who wasn’t killed (by Harry) with negation.

Fillmore (1968, p. 37) states that in the propositional core of a simple sentence, adjectives and nouns as well as verbs can function as predicates or predicators. The case hierarchy mentioned before is said to guide the operation of syntactic processes, in particular the selection of the surface subject in a sentence. The case which comes first in that list is made the subject. Fillmore (1968, p. 57) calls this transformational process of subject selection 'primary topicalization.' Topicalization, that is picking out a 'topic' or 'theme' of conversation, usually involves choice of a specific information 'focus,' which may be expressed by intonation and stress. The two choices are basically independent. 'Focus' is relevant for the distinction of new and old information, and denotes the foregrounding of certain elements of information. In the written language medium, where such a foregrounding by means of stress and intonation is impossible, either certain typographical devices are employed, or constructions such as for example the cleft sentence or pseudo-cleft sentence in English are used, which single out (topicalized) certain elements and thereby indicate a specific information focus. The same underlying conceptual structure consisting of a predicate and its arguments may appear on the surface in various shapes, as a full declarative sentence such as someone eats some apple, a pseudo-cleft sentence: what someone does is eat some apple, a nominalization: apple which can be eaten, or a compound: eating-apple. Topicalization may differ as to the result but also as concerns the elements which are made the topic. We may therefore distinguish between sentential topicalization and topicalization resulting in nominalizations or complex lexical items. Topicalization, that is the separation of elements and the selection of a 'topic' or 'theme,' is the
prerequisite for assigning 'focus' to the separated elements, that is semantic prominence and the distinction of old and new information. A single result of a topicalization process, for example a declarative sentence, may be realized with a wide range of focus assignment, depending on stress and intonation.

A variety of positions have been suggested to be the topic position in different languages. As for long distance topicalization, Chomsky (1977) argues that this movement obeys the same island conditions as wh-movement. This means that topicalization instantiates an A-bar movement to specifier of CP. However, the fact that topicalized constituents can appear after a complementizer indicates that there must be more than one phrasal projection on the top of the TP. Furthermore, evidence from Persian shows that more than one topic might undergo such A-bar movement.

\[
\text{Children seem to like the car.}
\]

3. Significance and Purpose of the Study

Topicalization and focalization are two processes in Persian whose constituents do not appear in canonical word order. In these constructions Topics and/or Focus phrases move from their base positions to Periphery positions sentence initially. These discourse motivated movements yield marked orders with specific readings. A number of analyses have been proposed to account for topicalization and/or focalization in Persian e.g. Karimi, 2005. Darzi, 2010. These analyses are based on Haegeman and Gueron’s split CP hypothesis (1999). However, these analyses suffer from the fact that they are not comprehensive in terms of considering all kinds of clauses (simple versus complex, independent versus dependent) and/or they do not account for sentences which
instantiate more than one topicalized and focalized constituents simultaneously. In this study we reevaluate the split CP hypothesis to find the extent to which it can appropriately account for Persian data.

4. Literature Review

4.1. Topicalization

Topic is an expression which represents old or familiar information which moved to the beginning of the clause through the process of topicalization. For instance, in the sentence "Shiraz, most tourists like to visit", the expression Shiraz is said to occupy topic position (Darzi, 2010). Topicalization and are a process in Persian whose constituents do not appear in canonical word order. In this construction Topics and/or Focus phrases move from their base positions to Periphery positions sentence initially. This discourse motivated movement yields marked orders with specific readings. A number of analyses have been proposed to account for topicalization in Persian e.g. Karimi, 2005. Darzi, 2010. These analyses are based on Haegeman and Gueron’s split CP hypothesis (1999). However, these analyses suffer from the fact that they are not comprehensive in terms of considering all kinds of clauses (simple versus complex, independent versus dependent) and/or they do not account for sentences which instantiate more than one topicalized constituents simultaneously. Topicalization in Persian occurs in simple clauses and complex clauses. Movement in complex clauses is licensed both to the periphery of the dependent clause or the matrix clause.

4.2. Theoretical Framework

A variety of positions have been suggested to be the topic position in different languages. As for long-distance topicalization, Chomsky (1977) argues that this movement obeys the same island
conditions as wh-movement. This means that topicalization instantiates an A-bar movement to specifier of CP. However, the fact that topicalized constituents can appear after a complementizer indicates that there must be more than one phrasal projection on the top of the TP. Furthermore, evidence from Persian shows that more than one topic might undergo such A-bar movement.

Most previous researches within the topicality paradigm have focused on subject agreement, with the recent exception of Dalrymple and Nikolaeva (2011). As for topicalization-origin accounts of object agreement, despite their continued popularity, there is a notable lack of actual empirical research in support of such a scenario. This study aims at revealing how English and Persian are different in topicalization.

4.3. Some Studies on Topicalization

In a very early discussion of the discourse constraints on Topicalization, Hankamer (1971) introduces a notion of 'presupposition', by which is meant the prior evocation in the discourse of the entity represented by the NP in question. His claim (p. 217) is that only 'presupposed NPs' may be proposed in Topicalization and from this it is inferred that indefinite NPs may not be preposed.

A somewhat different approach is taken in Kuno (1972), where the analysis is in terms of theme rather than presupposition. In particular, Kuno claims that the preposed NP in a Topicalization must represent the 'theme' (p. 310) and that a theme must be 'anaphoric' or generic (p. 301), where anaphoric is defined as representing an entity already 'in the registry of discourse', either because it has been mentioned previously or because it is in the 'permanent registry' (p. 271).
On the one hand, Davison (1984, p.814) claims that the only indefinite NPs that can occur felicitously in marked syntactic constructions, of which Topicalization is an exemplar, are specific indefinites.

4.4. Questions' Formation in Persian

Persian is a SOV language which indicates a free word order. Lotfi (2003) states that due to scrambling, which is a rudimentary phenomenon in Persian, in addition to the basic SOV word order; there are SVO, OSV, and OVS word orders. The following instances are from Lotfi (2003), which show the possible word orders in Persian:

(3a) Armin Elnaz-o did. (SOV)
Armin Elnaz-OM saw.
Armin saw Elnaz.
(b) Armin did Elnaz-o. (SVO)
Armin saw Elnaz-OM
Armin saw Elnaz.
(c) Elnaz-o Armin did. (OSV)
Elnaz-OM Armin saw
Armin saw Elnaz.
(d) Elnaz-o did Armin (OVS)
Elnaz-OM saw Armin.
Armin saw Elnaz.

As it can be seen from above, scrambling lets Persian speakers to afford different word orders which obeys movement limitations, such as Minimal Link Conditions (MLC) and are optional (Karimi, 1994). Persian contrary to English which shows objects by their position in the sentence uses an object marker (OM). This OM in Persian is *ra* which has several representations such as *–ro* and *–o* (as can be seen in the previous examples).
5. Methodology

The aim of this study was to compare the processing of topicalized sentences in L1 & L2 structures. This chapter clarifies the details of participants, the instruments for data collection, design and the procedures employed to answer question.

5.1. Participants

The participants of this study were 50 male and female EFL students at intermediate and advanced level of English language proficiency. All of participants were native speakers of Persian studying in Jahad Daneshgahi English language institute in Qazvin, Iran. The participants were administered two tests.

5.2. Instruments

In the present study, the following instruments were utilized:

1- Preliminary English language proficiency test (PET)
2- English Topicalization Test
3- Persian Topicalization Test

5.3. Data Collection Procedure

In order to achieve the aim of the study, the following procedures were followed. The data that is used in this study is the results of the three tests that were administered, that is, the Preliminary English Test (PET), the English topicalization test, and Persian topicalization test. First of all, PET test was administered to 73 participants to determine their English language proficiency level. Based on the results of PET, those 50 students whose scores were one standard deviation (9.39) plus and minus the mean (54.66) (scores between 45 and 64) were selected. The selected students were divided into two groups: Intermediate and Advanced. It consisted of 85 grammar,
vocabulary and reading comprehension questions. Throughout this study, the researcher sought to examine the possible effect that English and Persian topicalized structures may have on the ability to comprehend sentences. Then, to achieve this goal, the two groups took the same English and Persian topicalization test. We scored the test's answers separately to estimate the possible ability of EFL learners' comprehension. After the papers were scored, statistical procedures were being employed to measure the descriptive statistics of the results which will be reported. The results will then be used in verification of the research hypothesis.

5.4. Data Analysis

After all the tests were administered and the data were collected, to test the research hypotheses and to answer the research questions, The Wilcoxon Signed Rank Test was used to answer the research question of this study. We analyzed the obtained data to investigate the comprehension of topicalized structures by advanced and intermediate Iranian EFL learners. In addition we can find out whether having a high level of language proficiency will affect being aware of topicalized structures or not. In the case of having positive effect, the result enables the researcher to reject the null hypothesis.

6. Result and Discussions

The research question of this study aimed at exploring whether there is any difference between topicalization processes in English and Persian. In order to answer the research question of this study, before carrying out the inferential analysis, the test of normality was conducted to determine which test fit the data. If the data is normally distributed, the parametric test of paired-samples t-test would be used and it is confirmed that the data is not normally distributed, The Wilcoxon Signed Rank Test would be conducted. In the following, the test of normality is discussed for all of the data sets.
Table 6-1

Test of Normality on Argumentative Task and its Sub-components

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
</tr>
<tr>
<td>English Topicalization test</td>
<td>.222</td>
</tr>
<tr>
<td>Persian Topicalization test</td>
<td>.160</td>
</tr>
</tbody>
</table>

Instruction: If the Kolmogorov-Smirnov statistic is non-significant (Sig. value of more than .05) normality is met (Pallant, 2011). As seen in Table 6.2, the overall scores with probability values of 0.000 and 0.003 are not normally distributed. Thus, the Wilcoxon Signed Rank Test was conducted to compare the scores of the participants on the two occasions. In addition, the effect size for this test was calculated by dividing the z value by the square root of N (Pallant, 2011). The effect size was estimated using Cohen's (1988) criteria of .1 = small effect, .3 = medium effect, and .5 = large effect.

Table 6-2

Descriptive Statistics of the participants' scores on two occasions

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>25th</th>
<th>(Median)</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persian Topicalization test</td>
<td>50</td>
<td>7.8600</td>
<td>3.18164</td>
<td>2.00</td>
<td>20.00</td>
<td>5.7500</td>
<td>9.0000</td>
<td>10.000</td>
</tr>
</tbody>
</table>
As it can be seen in Table 6.2, the partial eta square index is .79, which shows that 79 percent of the variance in the comprehending scores is due to language type; this is quite a large effect size (.79 > .138). The attained results indicated that structure type (i.e., English and Persian topicalization process) influences comprehending significantly. Accordingly, the null hypothesis that states that there is no significant difference between topicalization process in English and Persian is rejected and we can claim that there is significant difference between topicalization processes in English and Persian. EFL learners processed the English topicalization test more better than Persian one. And in English test, the advanced group's performance was better than intermediate one.

Table 4-5

<table>
<thead>
<tr>
<th>Wilcoxon Rank Test Statisticsa</th>
<th>English Topicalization test - Persian Topicalization test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-6.162b</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.000</td>
</tr>
</tbody>
</table>

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

\[ r = \frac{z}{\sqrt{N}} = \frac{-6.162}{7.071} = 0.87 \]

Group Statistics
<table>
<thead>
<tr>
<th></th>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Topicalization</td>
<td>intermediate</td>
<td>25</td>
<td>16.600</td>
<td>6.36396</td>
<td>1.27279</td>
</tr>
<tr>
<td></td>
<td>advanced</td>
<td>25</td>
<td>24.400</td>
<td>2.97209</td>
<td>0.59442</td>
</tr>
<tr>
<td>Persian Topicalization</td>
<td>intermediate</td>
<td>25</td>
<td>6.1200</td>
<td>3.03205</td>
<td>0.60641</td>
</tr>
<tr>
<td></td>
<td>advanced</td>
<td>25</td>
<td>8.8400</td>
<td>2.89655</td>
<td>0.57931</td>
</tr>
</tbody>
</table>

Results: A Wilcoxon Signed Rank Test revealed a statistically significant reduction from the English Topicalization test scores to the Persian Topicalization test scores, \( z = -6.162, p < .0001 \), with a large effect size \( (r = .87) \). The median score on the Topicalization test decreased from English \( (Md = 24) \) to Persian \( (Md = 9) \).

The research question in the present study focused on the importance of processing topicalized sentences in English and Persian between students with advanced and intermediate levels. The result provides fairly strong support for the effect of topicalized structures on comprehension. Many studies have been conducted on topicalization process, but there are few studies on examining the difference between Persian and English topicalization process.

Concerning the topic under discussion, Birner and Mahootian (1996, pp.127-138) discuss the differences and similarities between discourse-functional constraints on inversion in English and the corresponding construction in Farsi. After offering different examples, they conclude that both languages allow a marked ordering of XSV. While this accounts for English topicalization, Farsi XSV corresponds to English inversion with regard to discourse functional constraints. In other words, XSV non-canonical word order represents English topicalization, whereas Farsi XSV and English XVS word order represent inversion. Further, findings of the present study are
compatible with the notion that in Farsi there is only one construction associated with two functions—inversion and topicalization—whereas in English two separate constructions can be observed (Mahootian, 2008, p.282).

In accordance with the result of the present study, Ward (1999) and Ward and Birner (2004) argue that postposed constructions preserve the old before new information structure paradigm by presenting relatively unfamiliar information in post verbal position.

In addition, the findings of the present study corroborate those of Lotfi(2003) who claims that Persian may exhibit both syntactic Wh- movement and Wh-in- Situ phenomena simultaneously. That is one can expect an optional movement in Persian, but Megerdoomian and Ganjavi(2001) argue against optional movement in Persian. They believe that it is not possible to apply an optional movement strategy to Persian and provide various types of evidence from distributional properties of the two constructions showing that topicalization in Persian Wh-in-situ and overt Wh-extraction are two distinct processes.

With regard to the topic under discussion, i.e. information structure and non-canonical word order, not much literature can be found in Farsi. Topicalization in Persian occurs in simple clauses and complex clauses. Movement in complex clauses is licensed both to the periphery of the dependent clause or the matrix clause (Rezai, Youhanaee, & Ghasemi, 2012). Some researchers such as Karimi (1994, p.69) and Mahootian (2008, p.281), while discussing preposing and topicalization, raise the issue briefly, but I could not find any independent study published on postposing in this language.

English and Persian show differences and similarities in their version constructions. First, at word-order level, the relative order of the XP and subject holds across languages: in all cases
the XP precedes the subject. However the resulting word order for English inversion is XVS and for Farsi is XSV. Although the XSV order of inversion, we see that in terms of pragmatic function both English and Farsi inversions abide by the same constraints. The information represented by the preposed XP must be at least as discourse familiar as the information represented by the subject. So in Farsi we see one construction associated with two functions where in English two separate constructions.

Two obvious possible explanations come to mind for these differences: First, the two languages differ in canonical word order, SVO for English and SOV for Farsi. Second, Farsi has somewhat richer morphology than English in particular the use of the particle-ra which can be used to mark topics.

Additionally, although non-canonical postposed constituents are alleged to represent discourse-new as well as hearer-new information, or at least information that is less familiar when compared with the constituents filling the preverbal position (Ward, 1999. Ward & Birner, 2004) in English and Italian, Farsi speakers impose the opposite constraints on the postposed constructions used. The NP constituents postposed were not found to represent new information to the hearer.

To sum up, Farsi speakers impose the opposite constraints on the postposed constructions used, the NP constituents postposed were not found to represent new information to the hearer. The data collected in spoken Farsi do not show similar results. Generally speaking, NPs were found to be triggered in post verbal position when the referent was hearer-old or hearer-old and discourse-old. None of the postposed NP constituents present hearer-new information. Farsi chooses an entity that is familiar to the discourse, the hearer or both.
These findings do not seem to support the distinction between internal interfaces and external interfaces in terms of L2 acquisition difficulty (see Belletti et al., 2007; Sorace, 2005, 2006, 2009; Sorace & Serratrice, 2002; Tsimpili and Sorace, 2006; and White, 2009). According to the proposal of the distinction, internal interfaces are relatively easier to acquire by L2 learners, but external interfaces, like syntax-discourse interfaces, are likely to be a locus of persistent difficulty in L2 acquisition. Given the findings in our study here, it seems reasonable to argue that external interfaces may not always be vulnerable in L2 acquisition and that they are acquirable, at least at advanced stages.

In sum, the acquisition of these properties in L2, an hypothesis such as Full Transfer / Full Access Hypothesis (Schwartz & Sprouse, 1994), predicts that it is possible to acquire all the properties of the L2 grammar, although, initially, there will be transfer of the L1 grammar. On the other hand, Sorace (2006), Robertson and Sorace (1999) and Tsimpili and Sorace (2006), for example, have argued that, unlike purely syntactic properties, the domain of interpretation remains "vulnerable" at the final stage of L2 acquisition. This asymmetry may be due to the fact that properties which involve an interface between two or more grammatical domains require more complex processing.

7. Conclusion

Results of the topicalization test revealed that the group with high level of English language proficiency (Advanced) had the best performance on processing English topicalized sentences. The other group with low level of English language proficiency (Intermediate) had the lowest mean and hardly processing topicalization in structure that are grammatically true. Based on the obtained results, the mean differences showed that there is a significant difference between the
English and Persian and also there is a significant difference between advanced and intermediate groups in processing topicalization. It means that the students of the advanced group in English test, (M = 24.4 and SD. = 0.59) had reached to the higher performance than those of the intermediate group in English test (M = 16.6 and SD. = 1.27). Then, the result of Persian test show that both groups cannot process Persian topicalized sentences well. The scores of the advanced group in Persian test with (M = 8.8 and SD. = 0.57) and those of the intermediate group in Persian test with (M = 6.12 and SD. = 0.606) indicated that there is no significant difference between them in processing Persian topicalization.

Therefore, the null hypothesis of this study suggested the topicalization process in English and Persian between both advanced and intermediate groups is not significantly different was rejected.

This study, like other studies, has some implications for different individuals including EFL teachers, researchers, and curriculum designers. The present study may also have implications for material writers, textbook authors, EFL package designers in that it may encourage them to be aware of topicalized input. Future researches can seek to answer several issues that are still not being investigated in this study. First, investigating the impact of instruction of topicalized sentences on comprehending and processing can be another area for research. Second, this study used multiple-choice questions to test learners’ reading comprehension. Other ways of testing reading comprehension may provide more precise results.

8. References


