Investigating the Effect of Translation Memory on English into Persian Translation

Abdul Amir Hazbavi

Department of English Language Translation and Teaching, Islamic Azad University, Bandar Abbas Branch, Iran

Abstract: A Translation Memory (TM) is a database in which a translator stores translations for future re-use. To examine the probable impact of TM tools on English into Persian translation, the present study was conducted. The subjects of the study, who were selected via a language proficiency test (TOEFL), were 90 Iranian undergraduate students of English translation from Bandar Abbas branches of Islamic Azad University and Payame Noor University, 45 as Control group and 45 as Experimental group. They were selected from among 120 B.A undergraduate students of English translation. To prepare homogenous groups of subjects and avoid bias, only the intermediate students were selected. The subjects first translated a text conventionally. Then during six training sessions, they got familiar with the concept of Translation Memory and learnt how to use the TM. Having passed the training sessions, the subjects took a translation test for which they were asked to translate a text of the same level of difficulty with that of the first text using the TM. Then, the results of the two tests were rated by 3 skilled translation raters according to the same translation checklist. Analyzing the scores of the translation tests using a T-test, the research hypothesis was supported which in turn indicated the positive effect of TM tools on English into Persian translation.

Key words: Machine Translation · Computer Assisted Translation · Translation Memory · Translation Database · MetaTexis

INTRODUCTION

Back Ground: The rising demand for translations over the last few decades has led to the recognition that software tools were urgently needed to help increase translators’ productivity and to support them in their efficient and effective delivery of accurate and consistent translations in ever-shorter time periods [1]. In the 1980s Machine Translation held great promises, but it has been steadily losing ground to Computer Assisted Translation (CAT) because the latter responds more realistically to actual needs. Research revived in the late 1980s because the dream of large-scale access to personal computers became a reality. Many computer companies started to work on CAT and it was then that the early word-processing programs were introduced [2].

In practice, computer-assisted translation, computer-aided translation, or CAT is a form of translation wherein a human translator translates texts using computer software designed to support and facilitates the process. This term is a broad and imprecise term covering a range of tools, from the fairly simple to the more complicated. Among CAT tools, Translation Memory (TM) is of a prominent role [3]. TM is defined by the Expert Advisory Group on Language Engineering Standards [4] as “a multilingual text archive containing (segmented, aligned, parsed and classified) multilingual texts, allowing storage and retrieval of aligned multilingual text segments against various search conditions. As stated by Webb, translation memory (also known as sentence memory) consists of a database that stores source and target language pairs of text segments that can be retrieved for use with present texts and texts to be translated in the future [5]. In a simple word, a Translation Memory is a database in which a translator stores translations for future re-use, either in the same text or other texts. Basically, the program records bilingual pairs: a source-language segment (usually a sentence) combined with a target-language segment. If an identical or similar source-language segment comes up later,
Research Questions and Hypothesis: On the basis of the information presented so far, it is assumed that using Translation Memory may have some influence on English into Persian translation.

Research Question: In accordance to the presented assumption, the following question is raised:

Does Translation Memory have any effect on English into Persian translation?

Null hypothesis: using Translation Memory has no effect on English into Persian translation.

Methodology
Participants: The subjects of the study, who were selected via language proficiency test (TOEFL), were 90 Iranian undergraduate students of English translation at Islamic Azad University, Bandar Abbas Branch and Payame Noor University, Bandar Abbas Branch. They were divided into two groups, 45 as Control group and 45 as Experimental group. To prepare homogenous groups of subjects and avoid bias, only the students who had intermediate performance at the proficiency test were selected. It should be noted that the age and gender of the participants was not controlled.

Instruments: Materials used in this study were a language proficiency test (TOEFL PbT), a pretest and a posttest with equal level of readability as well as a Translation Memory software named “MetaTexis” with a previously developed corpus which was used to conduct the experimental group posttest. It should be mentioned that MetaTexis is not a stand-alone-program rather it runs in Microsoft Word which means that all MetaTexis functions can be accessed through MS Word. The great advantage of the integration in MS Word was that the subject did not have to learn a completely new program. In fact they only had to learn some new functions while all functions of MS Word were available too. The version of MetaTexis used in this research was Pro 2.8. The corpus of the study contained 100 English into Persian segments which were the most frequent terms regarding computer science and was developed using Hezareh Dictionary which is considered to be the most authentic English into Persian Dictionary available. Besides, a translation checklist with numerical rating scale was used to rate all translations according to the same criteria.

Design: The present study is a pre-experimental design. The group under the study had never got training in any Computer Assisted Translation tool prior to the study. This group was divided into experimental and control...
group in which the experimental group received treatment that consisted of six sessions of training of how to use a Translation Memory.

**Procedure:** To start the research, the subjects were first selected via a language proficiency test. To make sure that the language proficiency test is a standard one, a Paper-based Test of English as a Foreign Language (TOEFL PbT) was provided. Those applicants, who had intermediate performance at the language proficiency test, were selected as the subjects of the study and were divided into two groups, 45 as experimental group and 45 as control group. The main section of the study began by conducting a pretest consisted of translating a text with controlled level of readability. The texts used in this study as pretest and posttest, were scientific texts related to computer science. To make sure that the second text used as posttest is not easier than the pretest and the difference observed is due to the impact of applying Translation Memory software to translation, the researcher had to make sure that our two tests - pretest and posttest- were of the same level of readability. In this study, MS Word was used to calculate the readability level of the pretest and posttests. The program calculated the following statistics for the two tests:

Having prepared the texts, the subjects in both groups were asked to translate the text in the conventional way as for pretest. Then, the subjects of the experimental group attended a six-session treatment aimed at introducing them with the concept of Translation Memory and teaching them how to work with Translation Memory software. Having finished the treatment, the subjects of both groups were provided with a text to translate, but this time the subjects of the experimental group were asked to translate the text with the help of the Translation Memory software.

**Data Analysis:** The results of the pretests and posttest were rated by 3 skilled translation raters according to the same translation checklist. Using SPSS version 16, the following analyses were made on subjects' scores resulted from above mentioned tests:

- Descriptive Statistics were run on the total scores of both groups

As shown in Table 2, the average of control group posttest (15.0673) has very slightly exceeded the average of control group pretest (14.9813). This increase is scientifically insignificant.

Table 3 presents different descriptive statistics of experimental group pretest and posttest. It shows remarkable increase between the average of pretest (15.6600) and the average of posttest (18.1600). This increase is statistically interpreted in later tables.

To make sure that the scores of all four tests have inter-rater reliability, the following set of correlations were run among scores of all the four tests

A correlation was run between Rater 1 and Rater 2 on the experimental group posttest scores

As shown on Table 4, the Correlation is significant at the 0.01 level (2-tailed); the r-value is 0.971, p $\leq$0.01 which means the scores of Rater 1 correlate significantly with the scores of Rater 2. Another point is that the p-value which represents the probability of observing an r-value of this size just by chance is 0.000.

A correlation was run between Rater 1 and Rater 3 on the control group pretest scores

Table 5 shows the Correlation between rater 1 and rater 3. The correlation is significant at the 0.01 level (2-tailed); the r-value is 0.959, p $\leq$0.01 which means the scores of Rater 1 correlate significantly with the scores of Rater 3.

A correlation was run between Rater 2 and Rater 3 on the control group posttest scores

The numbers in Table 6 indicate that the Correlation between rater 2 and rater 3 is significant at the 0.01 level (2-tailed); the r-value is 0.957, p $\leq$0.01 that means the scores of Rater 1 correlate significantly with the scores of Rater 3.

To assess the intra-rater reliability of our raters, they were asked to rate 15 translations that were selected randomly out of the total of 45 prior to the main rating. The following set of correlations were run among the first and the second scorings of each rater:

A correlation was run between the first and the second scores of rater 1 on experimental group pretest

Table 7 indicates that the Correlation between the first and the second scores of raters 1 is significant at the 0.01 level (2-tailed); the r-value for rater 1 is 0.978.

A correlation was run between the first and the second scores of rater 2 on experimental group pretest.

Table 8 shows that the Correlation between the first and the second scores of rater 2 is significant at the 0.01 level (2-tailed); the r-value for rater 2 is 0.972 and p $\leq$0.01.

A correlation was run between the first and the second scores of rater 3 on experimental group pretest.
Table 1: Pretest and Posttest Readability

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words</td>
<td>637</td>
<td>637</td>
</tr>
<tr>
<td>Characters</td>
<td>3260</td>
<td>3171</td>
</tr>
<tr>
<td>Paragraphs</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>Characters per Word</td>
<td>4.9</td>
<td>4.7</td>
</tr>
<tr>
<td>Passive sentences</td>
<td>20 percent</td>
<td>21 percent</td>
</tr>
<tr>
<td>Flesch Reading Ease</td>
<td>46.2</td>
<td>44.6</td>
</tr>
</tbody>
</table>

Table 2: Descriptive Statistics of control Group

<table>
<thead>
<tr>
<th></th>
<th>Control Group Pretest Average</th>
<th>Control Group Posttest Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>14.9813</td>
<td>15.0673</td>
</tr>
<tr>
<td>Median</td>
<td>15.1600</td>
<td>15.0000</td>
</tr>
<tr>
<td>Mode</td>
<td>14.16</td>
<td>15.00</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>1.58299</td>
<td>1.28914</td>
</tr>
<tr>
<td>Variance</td>
<td>2.506</td>
<td>1.662</td>
</tr>
<tr>
<td>Skewness</td>
<td>-.314</td>
<td>-.156</td>
</tr>
<tr>
<td>Std. Error of Skewness</td>
<td>.354</td>
<td>.354</td>
</tr>
<tr>
<td>Range</td>
<td>12.50</td>
<td>4.83</td>
</tr>
<tr>
<td>Minimum</td>
<td>11.66</td>
<td>12.50</td>
</tr>
<tr>
<td>Maximum</td>
<td>17.66</td>
<td>17.33</td>
</tr>
<tr>
<td>Sum</td>
<td>674.16</td>
<td>678.03</td>
</tr>
</tbody>
</table>

Table 3: Descriptive Statistics of Experimental Group

<table>
<thead>
<tr>
<th></th>
<th>Experimental Group Pretest Average</th>
<th>Experimental Group Posttest Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>15.2771</td>
<td>17.9367</td>
</tr>
<tr>
<td>Median</td>
<td>15.6600</td>
<td>18.1600</td>
</tr>
<tr>
<td>Mode</td>
<td>14.33</td>
<td>18.83</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>1.96109</td>
<td>1.12146</td>
</tr>
<tr>
<td>Variance</td>
<td>3.846</td>
<td>1.258</td>
</tr>
<tr>
<td>Skewness</td>
<td>-.399</td>
<td>-.922</td>
</tr>
<tr>
<td>Std. Error of Skewness</td>
<td>.354</td>
<td>.354</td>
</tr>
<tr>
<td>Range</td>
<td>5.00</td>
<td>14.66</td>
</tr>
<tr>
<td>Minimum</td>
<td>11.33</td>
<td>14.66</td>
</tr>
<tr>
<td>Maximum</td>
<td>18.33</td>
<td>19.66</td>
</tr>
<tr>
<td>Sum</td>
<td>687.47</td>
<td>807.15</td>
</tr>
</tbody>
</table>

Table 4: Rater 1 and Rater 2 Correlation

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater1 Pearson Correlation</td>
<td>1</td>
<td>.971**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>45</td>
</tr>
<tr>
<td>N</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>

Table 5: Rater 1 and Rater 3 Correlation

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater1 Pearson Correlation</td>
<td>1</td>
<td>.959**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>45</td>
</tr>
<tr>
<td>N</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
Table 6: Rater 2 and Rater 3 Correlation

<table>
<thead>
<tr>
<th></th>
<th>Rater2</th>
<th>Rater3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater2</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.957**</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>45</td>
</tr>
<tr>
<td>Rater3</td>
<td>Pearson Correlation</td>
<td>.957**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>45</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Table 7: The correlation between First and Second Scores of Rater 1

<table>
<thead>
<tr>
<th></th>
<th>First Scoring</th>
<th>Second Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Scoring</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.978**</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>15</td>
</tr>
<tr>
<td>Second Scoring</td>
<td>Pearson Correlation</td>
<td>.978**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>15</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Table 8: The correlation between First and Second Scores of Rater 2

<table>
<thead>
<tr>
<th></th>
<th>First Scoring</th>
<th>Second Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Scoring</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.972**</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>15</td>
</tr>
<tr>
<td>Second Scoring</td>
<td>Pearson Correlation</td>
<td>.972**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>15</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Table 9: The correlation between First and Second Scores of Rater 3

<table>
<thead>
<tr>
<th></th>
<th>First Scoring</th>
<th>Second Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Scoring</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.980**</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>15</td>
</tr>
<tr>
<td>Second Scoring</td>
<td>Pearson Correlation</td>
<td>.980**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>15</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Table 10: Paired Samples Test of Control Group

<table>
<thead>
<tr>
<th></th>
<th>Paired Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Pair 1 Control Group Pretest Average</td>
<td>Control Group Posttest Average</td>
</tr>
</tbody>
</table>

Table 11: Paired Samples Test of Experimental Group

<table>
<thead>
<tr>
<th></th>
<th>Paired Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Pair 1 Experimental Group Pretest Average</td>
<td>Experimental Group Posttest Average</td>
</tr>
</tbody>
</table>
The numbers in Table 9 show that the Correlation between the first and the second scores of rater 3 is significant at the 0.01 level (2-tailed); the r-value for rater 3 is 0.980 and $p < 0.01$. The numbers presented in the last three tables indicate that the first scores of all three raters correlate significantly with their second scores therefore; the level of intra-rater reliability is acceptable.

To see if there is any significant difference between the performance of our Control group subjects in pretest and posttest, a Paired Sample T-test was run among the averages of pretest and posttest.

The above table shows that the posttest average of Control group ($M=15.0673$) did not significantly exceeded pretest average of Control group ($M=14.9813$) while $t (44) = 0.349$ and Sig. (2-tailed) = 0.729 which is not $\leq 0.05$. This means that the difference observed among control group averages of pretest and posttest is not a significant reliable difference.

A Paired Sample T-test was run between the averages of pretest and posttest of the Experimental group to see if there is any significant difference among the averages of this group.

Table number 11 shows that the posttest average of Experimental group ($M=17.9367$) significantly exceeded pretest average of Experimental group ($M=15.2771$) while $t (44) = -15.481$ and Sig. (2-tailed) = 0.000 which is $< 0.05$. This means that the difference observed among pretest and posttest averages of experimental group is a significant difference.

**Conclusion and Implications:** In social sciences when the significance is below 0.05 (e.g. 0.000) the null hypothesis should be rejected and the alternative hypothesis will be accepted. As shown in section 5 of the data analysis (Table 11), the paired samples statistics of experimental group say that the t-test with 44 degrees of freedom was significant which in turn means that the null hypothesis can be rejected. This also means that the difference observed among Experimental group averages of pretest and posttest is a significant reliable difference. Ultimately, this indicates that Translation Memory tool had positive effect on English into Persian translation.

**Implications and Suggestions for Future Studies:** In the present research, beside the productivity improvement, it was scientifically approved that Translation Memory has positive effect on English into Persian translation and can result in translation quality improvement. The Iranian translation agencies as well as freelancers, who work on English into Persian translation, can take the advantages of integrating this technology into their own work in order to improve their productivity and to produce high quality translations in a shorter period of time. The use of Translation Memory tools can also turn them into an up-to-date translator equipped with the newest technologies on the market that in addition to earning more income has its own benefits too.

**REFERENCES**