

A Comparative Study of the Novel 'A Tale of Two Cities' and Its Persian Translation in Terms of Textual Cohesion: The Cases of Punctuation Marks, Sentencing and Paragraphing

Zahra Sojudifar

English Department, Fars Research and Science Branch, Islamic Azad University, Fars, Iran

Azadeh Nemati

Department of English Language Teaching, Jahrom branch, Islamic Azad University, Jahrom, Iran

Mohammad Reza Falahati Qadimi Fumani

Dep. of Computational Linguistics, Regional Information Center for Science and Technology (RICeST), Shiraz, Iran

Abstract—The present study aimed at finding possible differences in using punctuation marks between the novel 'A Tale of Two Cities' and its Persian translation and investigating whether TT sentences and paragraphs are as long as ST sentences and paragraphs. First, chapter one of the novel and its translation were selected. Then, punctuation marks including comma, full stop, semicolon, colon, hyphen, dash, and parenthesis in every thousand words were counted and the mean was calculated for each punctuation mark. Next, the number of words in the first 20 sentences and the first 10 paragraphs of each story were counted and the mean of sentence and paragraph length were calculated. Finally, T-Test and Wilcoxon Test were run. The results showed that there were significant differences between using comma, colon, semicolon, hyphen and dash in the ST and the TT. Moreover, TT sentences and paragraphs were as long as those of the ST.

Index Terms—cohesion, textual cohesion, punctuation, sentence, paragraph

I. INTRODUCTION

Textuality consists of seven features which must occur simultaneously to identify any oral or written extract as text. These seven criteria are: cohesion, coherence, intentionality, acceptability, informativity, situationality and intertextuality (formal or semantic connections with texts of the same type) (Beaugrande & Dressler 1981). Cohesion seems to be the only obligatory requisite of texture since it entails semantic and intertextual factors (Halliday & Hassan 1976). Cohesion enables us, by means of lexical, grammatical or other devices, to connect different items that make up a text (Baker 1992). A text has to be cohesive in continuation of statements or paragraphs (Tarnyikova 2009). Cohesion occurs where the interpretation of some elements in the text is dependent on that of another (Halliday & Hasan 1976).

Punctuation is among the factors which create cohesion in a text (Bernárdez 1982). Punctuation has two main functions: the marking of lexical, grammatical and rhetorical items and the mitigation of sentence or paragraph length (and hence complexity) (Quirk, Greenbaum, Leech & Svartvik 1985). Punctuation marks attach the sentences together and bring unity to the texts (Hyland 2005). The correct application of punctuation marks is notable for translators because of their vital role in transferring meaning from the Source Text (ST) to the Target Text (TT) (Kirkman 2006). In evaluation of a translated text to determine the degree to which the translator has managed to maintain discursive value intended by the ST, punctuation can offer critical guidelines. Punctuation needs to be observed especially in the translation of literary texts because the overall textual effect is the product of correct use of punctuation (Lotfipour Saedi 2001).

Translators often tend to automatically copy any graphic features of the ST to the TT (Ishenko 1998). Considering intratextual elements, most translators pay obsessive attention to the structure and lexemes of the text while minorities of them consider punctuation as an influential part in their work (Schwartz 2006).

The researcher observed the same problem (copying punctuation marks from the ST to the TT) in Persian translations of some English Literary texts. Translation of 'Little Women' by Fariba Dastom, translation of 'Great Expectations' and 'Jane Eyre' by Maryam Dastom are some examples.

A. Purpose of the Study

Regarding the above mentioned problems, the present study pursued the following objectives. The first objective of the study was to find possible differences in using punctuation marks between English novel 'A Tale of Two Cities' and

its Persian translation. The second objective of the study was to investigate whether TT sentences are as long as ST sentences. The third objective of the study was to investigate whether TT paragraphs are as long as ST paragraphs. Following these objectives, the following research questions were posed:

1. Are there any significant differences in using punctuation marks between English novel *'A Tale of Two Cities'* and its Persian translation?
2. Are TT sentences as long as ST sentences?
3. Are TT paragraphs as long as ST paragraphs?

B. Significance of the Study

Assessing three easily isolatable – but nevertheless frequently ignored – features of textual Cohesion (i.e. punctuation, sentencing, and paragraphing) is very important. Since assessments impact learning priorities in academic and professional settings, an assessment tool that focuses on these important, though often-overlooked, textual features, encourages novice translators to consider the target text globally, as a product involving a variety of features above and beyond lexis, for which they are professionally responsible (Baer & Bystrova-McIntyre 2009). The present study can check the practicality of Baer and Bystrova-McIntyre's (2009) frame work (See Section 2.5) between the English novel *'A Tale of Two Cities'* and its Persian translation. Moreover, most studies undertaken between English and Persian (Seddigh, Shokrpour, & Kafipour 2010; Rahimi & Ebrahimi 2012) dealt with lexical cohesion. Other studies (Øverås 1998; Querol 2009) investigated grammatical or lexical cohesion between English texts and languages other than Persian such as Norwegian and Spanish. But, the present study will focus on an area of cohesion (i.e textual cohesion) that has not been studied much before.

II. LITERATURE REVIEW

Cohesion is the textual quality which is responsible for making the sentences of a text seem to hang together (Morris & Hirst 1991). Cohesion is the property that distinguishes a sequence of sentences that form a discourse from a random sequence of sentences (Singh 1979). Speakers and writers often also provide internal cues as to how the parts of a text are linked together or how sentences are related to other sentences. These cues create cohesion in a text (Johnstone 2008). The identification of connections that are linguistically signaled, like those between a pronoun and a previous noun phrase, enables us to recognize the cohesion of a text (Widdowson 2007). Cohesive relationships within and between sentences create texture. Cohesive relationships have different types including reference, substitution, ellipsis, conjunction and lexical cohesion. The effect of special punctuation might be added to indicate a relationship between what has been said and what is about to be said. Observing these cohesive relationships will guarantee textual cohesion (Brown & Yule 1989).

Three easily isolatable – but nevertheless frequently ignored – features of textual cohesion are punctuation, sentencing, and paragraphing (Baer & Bystrova-McIntyre 2009). Punctuation is an essential aspect of discourse analysis, since it gives a semantic indication of the relationship between sentences and clauses, which may vary according to languages (Newmark 1988). Punctuation can be potent, but it is easily overlooked. So, translators are advised to make a separate comparative punctuation check on their version and the original (Newmark 1988). In what follows English and Persian punctuation marks will be introduced:

English major punctuation marks include: period, comma, semicolon, colon, dash, hyphen, and parentheses. Period (.) ends a sentence that makes a statement, direct command, or request. The period (full stop) is used after initials and many abbreviations (Wyrick 2008). Comma (,) separates two independent clauses joined by a coordinating conjunction. Coordinating conjunctions include “for”, “and”, “but”, “or”, “yet”, and “so”. The comma sets off nonessential phrases and clauses. Moreover, the comma sets off conjunctive adverbs such as “however”, “thus”, “consequently” and “therefore”. The comma is used in a series of words, phrases, or clauses, as well (Wyrick 2008). Semicolon (;) links two closely related independent clauses. The semicolon is used in a series between items that already contain internal punctuation (Wyrick 2008). Colon (:) is used to introduce a long or formal list. The colon is also used in the salutation of business or professional correspondence (Wyrick 2008). Dash (–) indicates a strong or sudden shift in thought. The dash is used before a statement that summarizes the proceeding thought (Wyrick 2008). Hyphen (-) joins words into a single adjective before a noun. The hyphen is used with some prefixes. The hyphen marks the separation of syllables when you divide a word at the end of a line (Wyrick 2008). And finally Parentheses () sets off words, dates, or statement that give additional information, explain, or qualify the main thought. Parentheses may also set off numbers in a list that appears within prose (Wyrick 2008).

Regarding Persian punctuation marks, Mohamadifar (2002) asserted, “the history of punctuation in Persian writing is not very old and it did not exist in classical writings. Its usage goes back to the advent of press industry in Iran” (p. 439). Period (.) is used after complete declarative sentences and after abbreviations (Yahaghi & Naseh 1996). Comma (,) shows pause and separates two successive words. The comma is used after and before appositive (Yahaghi & Naseh 1996). Semicolon (;) is used after related sentences and for separating independent meanings (Yahaghi & Naseh 1996). Colon (:) shows items that need definition or enumeration (Yahaghi & Naseh 1996). Parentheses () is used for giving additional information (Yahaghi & Naseh 1996). And finally, Hyphen (-) is used instead of “from-to” in dates (Yahaghi & Naseh 1996).

Regarding sentence and paragraph, sentence is the largest unit of grammatical organization within which parts of speech (e.g. noun, verbs and adverbs) and grammatical classes (e.g. word, phrase and clause) are said to function. In English a sentence normally contains one independent clause with a finite verb (Richards, Platt & Platt 1992). Paragraph is a unit of organization of written language which serves to indicate how the main ideas in a written text are grouped. In text linguistics, paragraphs are treated as indicators of the macro-structure of a text. They group sentences which belong together, generally, those which deal with the same topic. A new paragraph thus indicates a change in topic or sub-topic. (Richards et al. 1992).

Since the “textual turn” in Translation Studies, translation scholars and trainers have recognized global textual features, such as cohesion, to be of central importance for it is cohesion that creates “text” out of individual sentences (Neubert & Shreve 1992). Moreover, studies documenting translations done by novices and experts point to cohesion as a fundamental distinguishing trait. Because novices tend to translate at the level of word, phrase, and sentence, their translations often lack cohesion and so appear awkward and unfocused (Le 2004). This situation can in part be explained by the fact that the qualities that constitute cohesion are generally difficult to pinpoint and isolate. Considerable deficiencies in ‘discourse structure,’ i.e., in the way the sentences are combined into well-integrated paragraphs and these in turn into a well-constructed whole exist in translations which affect cohesion negatively (Baker 1992).

Baer and Bystrova-McIntyre (2009), in their study analyzed punctuation, sentencing, and paragraphing. In the study of punctuation, the researchers used two untagged comparable corpora of Russian and English editorials for the analysis. They selected the editorials randomly from leading daily Russian and American newspapers (*Izvestia* and *The New York Times*) regardless of their content. Each corpus consisted of 20,000 words (titles and names of the authors were not included in the word count). For the study of sentencing and paragraphing, the researchers used international news articles from the same newspapers, and contemporary literary texts in addition to editorials. They selected the articles and literary corpora (Tolstaya's *Perevodnye Kartinki* and Updike's *Seek My Face*) randomly, regardless of their content. The number of different texts in each category was 20 (i.e., 120 texts, totaling 116,140 words). The results of the study showed that use of commas, colons, dashes and parentheses in the Russian editorials occurred with significantly greater frequency than in the English editorials, while the use of semicolons and hyphens were not significantly different. Significant differences in the use of commas, colons, dashes, and parentheses in English and Russian implied different grammatical and stylistic principles underlying the use of punctuation in those languages. The results also showed that the average number of words per sentence was significantly higher for English for all three text types. For paragraphs, the words-per-paragraph count was significantly higher for Russian international news. English international news reports revealed a tendency for more concise, focused paragraphs, often 1–2 sentences long. For literary texts, the result was not statistically significant (e.g., Updike, 728 words per paragraph).

III. METHODOLOGY

The present study used a mixture of qualitative and quantitative methods in collecting data, registering the observed cases, and analyzing the overall characteristics of sentencing, paragraphing and punctuation usage in both English and Persian languages. So, the design used in the present study is a descriptive-comparative one.

Two main sources were used as the materials in the present study: “*A Tale of Two Cities*” and its Persian translation, ‘*Da'asta'an-e dou Sha'hr*’, by Ibrahim Unesi. The first chapter of the novel, ‘*Recalled to Life*’, and its Persian translation were selected purposely as the sample in the present study. The reason behind selecting the sample was that it was the shortest chapter of the book. Had the researcher selected the other chapters it would not have been possible to complete the project within the expected time due to big data size. Therefore, the sample includes stories the period, the mail, the night shadows, the preparation, the wine-shop, the shoemaker in the ST and their translation in the TT.

Procedure of Data Collection and Analysis

To undertake the study and having selected the sample data, punctuation marks (i.e. coma, full stop, semicolon, colon, hyphen, dash, and parenthesis) were counted in every 1000 words (using word count option in Microsoft Word). Frequency and mean scores were computed here. Then, the number of words in the first 20 sentences of each story was counted using word count option in Microsoft Word. Frequency and mean scores were also computed. Later, the number of words in the first 10 paragraphs of each story was computed. After that, Performing Kolmogorov-Smirnov Test was used to check the normality of variables. Finally, paired sample T-Test and the Wilcoxon Test were used to analyze the data.

IV. DATA ANALYSIS AND RESULTS

This part presents and analyzes the results of different procedures done in conducting the study.

A. Analysis of Results of Mean Number of Punctuations

Punctuations in every thousand words were counted and registered in the ST and the TT (See Appendix A). So the ST had 17 groups of thousand words (roughly 17000 words) and the TT had 19 thousand words.

TABLE 4.1
STATISTICS OF NUMBER OF PUNCTUATIONS IN THE ST AND THE TT

Punctuation	Statistics		N	Mean	Std. Deviation	Minimum	Maximum
	Source						
Comma	English (ST)		17	98.29	7.880	82	111
	Persian (TT)		19	40.05	9.083	25	63
Fullstop	English		17	44.65	12.145	21	63
	Persian		19	43.53	16.514	24	95
Semicolon	English		17	7.71	4.552	2	19
	Persian		19	2.58	2.090	0	7
Colon	English		17	2.06	1.435	0	4
	Persian		19	8.11	4.280	1	17
Hyphen	English		17	8.35	3.757	0	15
	Persian		19	1.74	1.368	0	4
Dash	English		17	5.47	4.303	1	17
	Persian		19	0.00	0.000	0	0
Parenthesis	English		17	1.29	1.359	0	5
	Persian		19	0.53	0.697	0	2

As indicated in Table 4.1, mean number of comma was 98.29 and 40.05 respectively in the ST and the TT. Mean number of full stop was 44.65 and 43.53 respectively in the ST and the TT. Mean number of semicolon was 7.71 and 4.58 respectively in the ST and the TT. Mean number of colon in the ST was 2.06 and 8.11 respectively in the ST and the TT. Mean number of hyphen was 8.35 and 1.74 respectively in the ST and the TT. Mean number of dash was 5.47 and 0.00 respectively in the ST and the TT. Mean number of parenthesis was 1.29 and 0.53 respectively in the ST and the TT. Thus, comma, semicolon, hyphen and parenthesis were used more frequently in the ST than the TT. Colon was used more frequently in the TT than the ST. Full stop was used equally in the ST and the TT and dash was used only in the ST, at all.

B. Analysis of Results of Mean Number of Words per Sentence

To collect data, number of words in first 20 sentences of stories of chapter one of the novel in the ST and the TT were counted and registered (See Appendix B).

TABLE 4.2
STATISTICS OF NUMBER OF WORDS PER SENTENCE IN THE ST AND THE TT

Story	Statistics		No. of sentences	Mean	Std. Deviation	Minimum	Maximum
	Source						
1	English (ST)		19	52.89	48.189	13	222
	Persian (TT)		20	33.70	25.795	6	94
2	English		20	32.40	21.402	4	79
	Persian		20	43.35	28.943	5	137
3	English		20	32.20	17.213	6	72
	Persian		20	51.25	30.409	11	117
4	English		20	26.25	18.467	4	67
	Persian		20	32.45	23.415	5	77
5	English		20	40.75	22.311	10	97
	Persian		20	36.70	22.734	10	101
6	English		20	21.75	10.249	9	45
	Persian		20	20.65	16.671	4	67
Total	English		119	34.22	27.008	4	222
	Persian		120	36.35	26.321	4	137

As Table 4.2 illustrates, mean number of words per sentence in the ST was 34.22 with standard deviation of 27.008, minimum number of 4 and maximum number of 222. In the TT, mean number of words per sentence was 36.35 with standard deviation of 26.321, minimum number of 4 and maximum number of 137.

C. Analysis of Results of Mean Number of Words per Paragraph

To collect data, number of words in first 10 paragraphs of stories of chapter one of the novel in the ST and the TT were counted and registered (See Appendix C).

TABLE 4.3
STATISTICS OF NUMBER OF WORDS PER PARAGRAPH IN THE ST AND THE TT

Story	Statistics		No. of paragraphs	Mean	Std. Deviation	Minimum	Maximum
	Source						
1	English (ST)		6	167.00	97.759	66	318
	Persian (TT)		6	162.00	103.317	45	318
2	English		10	70.00	72.210	4	209
	Persian		10	91.10	86.765	3	250
3	English		10	111.30	73.678	38	278
	Persian		10	111.50	92.257	16	303
4	English		10	69.50	72.489	9	251
	Persian		10	79.50	86.282	10	296
5	English		10	145.30	113.315	20	325
	Persian		10	170.80	146.718	20	405
6	English		10	33.50	39.190	2	137
	Persian		10	31.30	51.394	1	168
Total	English		56	94.61	88.407	2	325
	Persian		56	103.82	105.151	1	405

As Table 4.3 shows, mean number of words per paragraph in the ST was 94.61 with standard deviation of 88.407, minimum number of 2 and maximum number of 325. In the TT, mean number of words per paragraph was 103.82 with standard deviation of 105.151, minimum number of 1 and maximum number of 405.

D. Analysis of Results of Kolmogorov-Smirnov Test

To test the research hypotheses either t-test or Wilcoxon test was needed. The t-test is more precise, but needs normal distribution of variables as prerequisite. Thus, first normality of the variables distribution was investigated by Kolmogorov-Smirnov Test.

TABLE 4.4
KOLMOGOROV-SMIRNOV TEST OF NORMALITY

Punctuation	Statistics		Z	Sig. (p)
	Source			
Comma	English (ST)		0.717	0.683
	Persian (TT)		0.478	0.976
Full stop	English		0.654	0.786
	Persian		0.796	0.550
Semicolon	English		0.860	0.450
	Persian		0.625	0.829
Colon	English		1.128	0.157
	Persian		0.516	0.952
Hyphen	English		0.610	0.851
	Persian		0.779	0.579
Dash	English		0.915	0.372
	Persian			
Parenthesis	English		0.759	0.612
	Persian		1.543	0.017
Word per paragraph	English		1.217	0.104
	Persian		1.312	0.064
Word per sentence	English		1.598	0.012
	Persian		1.247	0.089

In Table 4.4, the p-values were greater than 0.05 (Sig.>0.05) for variables including commas (in the ST and the TT), full stop (in the ST and the TT), semicolon (in the ST and the TT), colon (in the ST and the TT), hyphen (in the ST and the TT), dash (in the ST), parenthesis (in the ST), word per sentence (in the TT), and word per paragraph (in the ST and the TT). So, for these variables the statistics were not significant. This means that the distributions were normal. For variables dash (in the TT), parenthesis (in the TT) and word per sentence (in the ST) the statistics were significant (Sig.<0.05), so these variables deviated from normal distribution.

E. Analysis of Results of Paired T-tests

The paired t-test was conducted for those variables with normal distributions including commas, full stop, semicolon, colon, hyphen, and word per paragraph. Data considered to be paired (i.e. any English data collected was accompanied by its Persian translation).

TABLE 4.5
 PAIRED T-TEST FOR COMPARISON OF PUNCTUATIONS IN THE ST AND THE TT

Variable	Statistics		N	Mean	Std. Deviation	Std. Error Mean	Mean Difference	t	Df	Sig. (p)
	Source									
Comma	English (ST)		17	98.29	7.880	1.911	60.059	22.600	16	0.001
	Persian (TT)		17	38.24	7.242	1.756				
Full stop	English		17	44.65	12.145	2.946	1.647	0.376	16	0.712
	Persian		17	43.00	17.346	4.207				
Semicolon	English		17	7.71	4.552	1.104	4.941	4.080	16	0.001
	Persian		17	2.76	2.107	0.511				
Colon	English		17	2.06	1.435	0.348	-5.882	-4.991	16	0.001
	Persian		17	7.94	4.337	1.052				
Hyphen	English		17	8.35	3.757	0.911	6.471	6.885	16	0.001
	Persian		17	1.88	1.364	0.331				

Table 4.5 indicates that mean number of comma, semicolon and hyphen were more frequent in the ST than in the TT, while colon was more frequent in the TT than in the ST. The result of t-test was not significant for number of full stops ($p=0.712>0.05$).

TABLE 4.6
 PAIRED T-TEST FOR COMPARISON OF PARAGRAPH LENGTH IN THE ST AND THE TT

Variable	Statistics		N	Mean	Std. Deviation	Std. Error Mean	Mean Difference	t	Df	Sig. (p)
	Source									
Word per paragraph	English		56	94.61	88.407	11.814	-9.214	-1.356	55	0.181
	Persian		56	103.82	105.151	14.051				

The result of the test was not significant for number of words per paragraph ($p=0.181>0.05$).

F. Analysis of Results of Wilcoxon Test

The Wilcoxon test was conducted for variables which failed to have normal distribution including dash, parenthesis, and word per sentence. Data considered to be paired (i.e. any English data collected was accompanied by its Persian translation).

TABLE 4.7
 THE WILCOXON-TEST FOR COMPARISON OF PUNCTUATIONS IN THE ST AND THE TT

		N	Mean Rank	Sum of Ranks	Z	Sig. (p)
Dash	Negative Ranks	17 ^a	9.00	153.00	-3.630	0.001
	Positive Ranks	0 ^b	0.00	0.00		
	Ties	0 ^c				
	Total	17				
Parenthesis	Negative Ranks	10 ^a	7.70	77.00	-1.568	0.117
	Positive Ranks	4 ^b	7.00	28.00		
	Ties	3 ^c				
	Total	17				

a. Persian<English, b. Persian>English, c. Persian=English

Table 4.7 illustrates that the test statistics was significant for number of dashes ($p=0.001<0.05$), but it was not significant for number of parentheses ($p=0.117>0.05$).

TABLE 4.8
 THE WILCOXON-TEST FOR COMPARISON OF SENTENCE LENGTH IN THE ST AND THE TT

		N	Mean Rank	Sum of Ranks	Z	Sig. (p)
words per sentences	Negative Ranks	58 ^a	54.84	3181.00	-0.736	0.462
	Positive Ranks	59 ^b	63.08	3722.00		
	Ties	2 ^c				
	Total	119				

a. Persian<English, b. Persian>English, c. Persian=English

As indicated in Table 4.8, the test statistics was not significant ($p=0.462>0.05$). Therefore, the number of words per sentence was not significantly different in the ST and the TT.

V. DISCUSSION

In this part, each research question and its relevant findings will be discussed:

The first research question of the study was, “Are there any significant differences in using punctuation marks between English novel ‘*A Tale of Two Cities*’ and its Persian translation?”. To answer the this research question, punctuation marks including comma, full stop, semicolon, colon, hyphen, dash, and parenthesis in every thousand

words were counted in the ST and the TT and their frequency was registered (See Appendix A). Then, the mean number of each punctuation mark was calculated both in the ST and in the TT. As indicated in Table 4.1, the application of comma in the ST (mean= 98.29) was almost two times than in the TT (mean= 40.05), full stops in the ST (mean= 44.65) and the TT (mean= 43.53) were used equally, semicolon was used almost three times in the ST (mean=7.71) than in the TT (mean=2.58), colon was used almost four times in the TT (mean=8.11) than in the ST (mean=2.06), hyphen was used almost five times in the ST (mean= 8.35) than in the TT (mean= 1.74), dash was not used in the TT at all, while the mean number of dash in the ST was 5.47, and finally the frequency of appearance of parenthesis was almost two times in the ST (mean= 1.29) than in the TT (mean=0.53). Considering the above mentioned means, it could be inferred that comma, semicolon, hyphen, dash and parenthesis were used more frequently in the ST than in the TT while colon was used more in the TT than in the ST.

To make sure that the above mentioned differences were significant or not, a paired sample t-test was conducted for punctuation marks with normal distribution including comma, colon, semicolon, hyphen and full stop. The Wilcoxon test was also conducted for punctuation marks which failed to have normal distribution including dash and parenthesis. A Kolmogorov-Smirnov test was run in advance to check the normality of variables' distribution (see Table 4.4). As Table 4.5 indicated, the number of commas, semicolons, colons and hyphens were significantly different between the ST and the TT ($p=0.001 < 0.05$). But the number of full stops was not significantly different between the ST and the TT ($p=0.712 > 0.05$). Thus, it was inferred that the number of full stop was statistically the same in the ST and the TT. As Table 4.7 showed, the number of dash was significantly different between the ST and the TT ($p=0.001 < 0.05$). But the number of parenthesis was not significantly different between the ST and the TT ($p=0.117 > 0.05$). Therefore, the first research hypothesis (H_01 : There is no significant difference in using punctuation marks between English novel 'A Tale of Two Cities' and its Persian translation) was rejected in terms of comma, colon, semicolon, and dash. But it was accepted in terms of full stop and parenthesis.

The second research question of the study was, "Are TT sentences as long as ST sentences?". To answer the research question, the initial 20 sentences of each story were selected in the ST and the TT, and the number of words in each sentence was counted and registered (See Appendix B). Then, mean of sentence length in each story was calculated both in the ST and in the TT. As Table 4.2 showed, mean number of words per sentence in the ST was 34.22 and that in the TT was 36.35. Regarding mentioned mean numbers in the ST and the TT, it seems that TT sentences were as long as ST sentences. Further, a Wilcoxon test was performed. The result of test shown in Table 4.8 indicated that the test statistics was not significant ($p=0.462 > 0.05$). That means the number of words per sentences were not significantly different between the ST and the TT. In other words, TT sentences were found to be as long as ST sentences. Thus, the second research hypothesis (H_02 : Target text sentences are not as long as source text sentences) was rejected.

The third research question was, "Are TT paragraphs as long as ST paragraphs?". To answer this research question, the initial 10 paragraphs of each story were selected in the ST and the TT and the number of words in each paragraph was counted and registered (See Appendix C). Then, mean of paragraph length in each story was calculated both in the ST and in the TT. As Table 4.3 illustrated, mean number of words per paragraph in the ST was 94.61 and that in the TT was 103.82. Therefore, TT paragraphs were found to be as long as ST paragraphs.

Further, a paired sample t-test was performed. The results of the t-test (Table 4.6) was not significant for the number of words per paragraph ($p=0.181 > 0.05$). That means the number of words per paragraph was not significantly different between the ST and the TT. In other words, TT paragraphs were as long as ST paragraphs. Consequently, the third research hypothesis (H_03 : Target text paragraphs were not as long as source text paragraphs) was rejected.

Conclusions

The present study pursued three objectives: (1) finding possible differences in using punctuation marks between English novel 'A Tale of Two Cities' and its Persian translation, (2) investigating whether TT sentences were as long as ST sentences, and (3) investigating whether TT paragraphs were as long as ST paragraphs. In line with the mentioned objectives, the researcher conducted a comparative study between the stories of the first chapter of the novel 'A Tale of Two Cities' and its Persian translations. The researcher collected necessary data, analyzed it, and discussed the obtained results.

Based on the discussion of the results, the following conclusions were drawn: (1) there were significant differences between using comma, colon, semicolon, hyphen and dash in the ST and the TT; comma, semicolon, hyphen and dash were used more frequently in the ST while colon was used more frequently in the TT, (2) there were not any significant differences between using other punctuation marks (i.e. full stop and parenthesis) in the ST and the TT, (3) TT sentences were found to be as long as ST sentences, and (4) TT paragraphs were found to be as long as ST paragraphs. Moreover, it was found that Baer and Bystrova-McIntyre's (2009) frame work (section 2.5) could be used between English and Persian Languages. Although the frame work, in the present thesis, delivered different results, it supported Baer and Bystrova-McIntyre's finding (i.e. differences in using punctuation marks between compared languages imply different grammatical and stylistic principle in those languages).

For the findings of the present study the following supportive statements on textual cohesion, were observed:

Every language has its own battery of certain cohesive devices for creating links between textual elements and there are different devices in different languages for achieving cohesive effects (Xi 2010). While every language has at its disposal a set of devices for maintaining textual cohesion, different languages have preferences for certain of these

devices and neglect certain others (James 1980). In addition to the fact that each language has general preferences for certain cohesive devices, it also has specific preferences for certain cohesive devices that are sensitive to text type (e.g. literary texts) (Xu 1996). There are specific preferences for certain cohesive devices in literary texts (Halliday & Hassan 1976). A text should be coherent with respect to itself, and therefore it is cohesive (Halliday & Hassan 1976).

To conclude, each language has its own rules of constructing cohesion; however, there are some similarities between the compared languages. When the compared languages are similar, the reconstruction of textual cohesion seems to be easier. As mentioned earlier (in section 5.2), each text should be cohesive with respect to itself; therefore, a translation considered as a text should be cohesive, too. Differences between languages are important issues in translation. Focusing on them will help translators to present a correct translation in terms of punctuation marks. Considering factors such as paragraph or sentence length in different languages will also pave the way in rendering an adequate translation.

APPENDIX A. PUNCTUATION MARKS IN EVERY THOUSAND WORDS IN THE ST AND THE TT

	Text	No of Words	Comma	Full stop	Semicolon	Colon	Hyphen	Dash	Parenthesis
1 st thousand words	English (ST)	1007	94	21	12	3	9	3	1
	Persian (TT)	1010	40	25	4	3	1	0	0
2 nd thousand words	English (ST)	1002	90	35	7	0	14	2	0
	Persian (TT)	1014	35	29	0	13	4	0	0
3 rd thousand words	English (ST)	1017	97	63	3	1	8	2	2
	Persian (TT)	1000	29	48	3	17	4	0	1
4 th thousand words	English (ST)	1011	102	28	5	0	11	4	2
	Persian (TT)	1002	45	24	6	2	2	0	0
5 th thousand words	English (ST)	1024	101	47	4	3	10	7	1
	Persian (TT)	1002	25	31	2	9	2	0	1
6 th thousand words	English (ST)	1013	94	50	2	3	6	2	0
	Persian (TT)	1004	49	45	3	1	0	0	1
7 th thousand words	English (ST)	1004	102	46	6	1	7	11	2
	Persian (TT)	1007	36	37	0	6	1	0	0
8 th thousand words	English (ST)	1027	82	48	14	2	8	17	3
	Persian (TT)	1041	43	95	3	6	4	0	0
9 th thousand words	English (ST)	1021	104	45	11	4	6	5	5
	Persian (TT)	1011	42	57	4	4	3	0	1
10 th thousand words	English (ST)	1013	100	26	19	0	15	3	0
	Persian (TT)	1005	49	46	1	9	0	0	1
11 th thousand words	English (ST)	1032	106	39	9	3	11	4	2
	Persian (TT)	1003	27	24	0	5	2	0	0
12 th thousand words	English (ST)	1008	103	54	5	0	7	4	0
	Persian (TT)	1034	46	31	5	10	1	0	2
13 th thousand words	English (ST)	1012	104	61	6	3	11	6	0
	Persian (TT)	1025	37	44	1	14	2	0	0
14 th thousand words	English (ST)	1037	85	56	6	4	8	1	2
	Persian (TT)	999	31	43	1	10	3	0	1
15 th thousand words	English (ST)	1027	91	52	7	3	2	2	1
	Persian (TT)	1002	40	45	4	10	2	0	0
16 th thousand words	English (ST)	1000	105	36	3	2	0	11	0
	Persian (TT)	997	37	60	3	9	1	0	2
17 th thousand words	English (ST)	1128	111	52	12	3	9	9	1
	Persian (TT)	999	39	47	7	7	0	0	0
18 th thousand words	English (ST)	-	-	-	-	-	-	-	-
	Persian (TT)	1004	63	43	0	6	1	0	0
19 th thousand words	English (ST)	-	-	-	-	-	-	-	-
	Persian (TT)	1364	48	53	2	13	0	0	0

APPENDIX B. NUMBER OF WORDS IN SENTENCE IN THE ST AND THE TT

	Text	Story 1	Story 2	Story 3	Story 4	Story 5	Story 6
Number of words in sentence 1	English (ST)	118	26	22	28	13	17
	Persian (TT)	6	39	20	57	10	20
Number of words in sentence 2	English (ST)	40	17	72	25	40	19
	Persian (TT)	6	137	86	32	37	10
Number of words in sentence 3	English (ST)	27	79	23	25	21	18
	Persian (TT)	7	48	11	77	21	20
Number of words in sentence 4	English (ST)	13	45	30	24	46	9
	Persian (TT)	6	62	58	9	71	11
Number of words in sentence 5	English (ST)	13	30	23	31	35	20
	Persian (TT)	7	26	85	14	51	60
Number of words in sentence 6	English (ST)	37	44	29	16	88	13
	Persian (TT)	62	29	93	13	28	67
Number of words in sentence 7	English (ST)	35	19	34	6	51	15
	Persian (TT)	45	33	50	8	101	4
Number of words in sentence 8	English (ST)	51	27	53	6	24	30
	Persian (TT)	6	45	42	5	10	11
Number of words in sentence 9	English (ST)	30	32	36	9	10	13
	Persian (TT)	16	16	36	17	42	10
Number of words in sentence 10	English (ST)	68	43	46	6	48	37
	Persian (TT)	13	16	37	7	23	35
Number of words in sentence 11	English (ST)	55	17	30	8	31	45
	Persian (TT)	38	30	13	14	22	13
Number of words in sentence 12	English (ST)	58	14	32	4	97	21
	Persian (TT)	37	52	20	64	13	17
Number of words in sentence 13	English (ST)	42	46	13	62	26	36
	Persian (TT)	44	68	89	75	44	23
Number of words in sentence 14	English (ST)	16	28	6	67	17	9
	Persian (TT)	28	66	22	15	37	14
Number of words in sentence 15	English (ST)	222	33	23	13	40	16
	Persian (TT)	71	36	22	29	52	32
Number of words in sentence 16	English (ST)	81	71	28	34	52	24
	Persian (TT)	57	5	117	51	47	10
Number of words in sentence 17	English (ST)	25	64	44	41	27	17
	Persian (TT)	94	67	42	47	18	8
Number of words in sentence 18	English (ST)	41	4	68	36	50	18
	Persian (TT)	31	41	59	27	59	8
Number of words in sentence 19	English (ST)	33	4	14	43	53	39
	Persian (TT)	60	42	56	42	24	19
Number of words in sentence 20	English (ST)	—	5	18	41	46	19
	Persian (TT)	40	9	67	46	24	21

APPENDIX C. NUMBER OF WORDS IN PARAGRAPH IN THE ST AND THE TT

	Text	Story 1	Story 2	Story 3	Story 4	Story 5	Story 6
Number of words in paragraph 1	English (ST)	118	167	278	53	53	17
	Persian (TT)	93	179	303	55	47	16
Number of words in paragraph 2	English (ST)	66	93	87	80	241	22
	Persian (TT)	45	139	93	111	326	1
Number of words in paragraph 3	English (ST)	148	101	144	9	210	2
	Persian (TT)	154	108	165	10	239	19
Number of words in paragraph 4	English (ST)	253	209	42	27	135	8
	Persian (TT)	251	250	16	29	138	1
Number of words in paragraph 5	English (ST)	318	64	95	16	27	37
	Persian (TT)	318	38	25	13	24	6
Number of words in paragraph 6	English (ST)	99	31	82	45	292	137
	Persian (TT)	111	4	108	55	351	13
Number of words in paragraph 7	English (ST)	—	4	38	129	325	45
	Persian (TT)	—	3	85	139	405	5
Number of words in paragraph 8	English (ST)	—	7	183	47	85	27
	Persian (TT)	—	5	44	44	87	20
Number of words in paragraph 9	English (ST)	—	5	57	251	65	36
	Persian (TT)	—	150	220	296	71	168
Number of words in paragraph 10	English (ST)	—	19	107	38	20	4
	Persian (TT)	—	35	56	43	20	64

REFERENCES

- [1] Baer, B. J., & Mc-Intyre, T. (2009). Assessing cohesion: Developing assessment tools on the basis of comparable corpora. In C.V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 159-183). Philadelphia: John Benjamins.
- [2] Baker, M. (1992). *In other words*. NY: Rutledge.
- [3] Beaugrande, R., & Dressler, W. U. (1981). *Introduction to Text Linguistics*. London: Longman.
- [4] Bern ádez, E. (1982). *Introducción a la Lingüística del Texto*. Madrid: Espasa-Calpe.
- [5] Brown, G., & Yule, G. (1989). *Discourse Analysis*. Cambridge: CUP.
- [6] Callow, K. (1974). *Discourse Considerations in Translating the Word of God*. Michigan: Zondervan.
- [7] Dickens, Ch. (1985). *A Tale of Two Cities*. Philadelphia: T.B. Peterson and Brothers.
- [8] Halliday, M. A. K., & Hassan, R. (1976). *Cohesion in English*. London: Longman.
- [9] Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- [10] Ishenko, M. (1998). Translating punctuation marks: punctuating and formatting issues in English-Russian translation. *Proceedings of the 39th Annual Conference of the American Translators Association*, pp. 155–174. Retrieved December 26, 2013, from <https://linguistlist.org/issues>.
- [11] James, C. (1980). *Contrastive Analysis*. London: Longman.
- [12] Johnstone, B. (2008). *Discourse Analysis*. (2nd ed.). UK: Blackwell Publishing.
- [13] Kachroo, B. (1980). Textual cohesion and translation. Retrieved December 28, 2013, from <http://jaits.sakura.ne.jp>.
- [14] Kirkman, J. (2006). *Punctuation Matters: Advice on Punctuation for Scientific and Technical Writing* (4th ed.). NY: Routledge.
- [15] Le, E. (2004). The role of paragraphs in the construction of coherence – text linguistics and translation studies. *International Review of Applied Linguistics in Language Teaching*, 42, 259–275. Retrieved January 4, 2014, from <http://eric.ed.gov>.
- [16] Lotfipour Saedi, K. (2001). Lexical cohesion and translation equivalence. Retrieved June, 28, 2014, from <http://www.neiu.edu/~circill/.../lexical>.
- [17] Mohamadifar, M. (2002). *Punctuation Guide*. Tehran: Diba.
- [18] Morris, J., & Hirst, G. (1991). Lexical cohesion computed by Thesaural relations as an indicator of the structure of text. *The Journal of Computational Linguistics*, 17, 21-48. Retrieved June 17, 2014, from <http://dl.acm.org/citation.cfm>.
- [19] Neubert, A., & Shreve, G. M. (1992). *Translation as Text*. Kent: Kent State University.
- [20] Newmark, P. (1987). *The Use of Systemic Linguistics in Translation Analysis and Criticism*. Philadelphia: John Benjamins.
- [21] Newmark, P. (1988). *A Textbook of Translation*. London: Prentice Hall.
- [22] Øverås, L. (1998). In search of the third code: an investigation of norms in literary translation. *Meta*, 43, 571-88. Retrieved July 6, 2014, from <http://nelson.cen.umontreal.ca>.
- [23] Querol, M. (2009). Substitution as a device of grammatical cohesion in English narratives and its translation into Spanish. *Jornades de Foment de la Investigació* Retrieved January 5, 2014, from <http://www.uji.es>.
- [24] Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- [25] Rahimi, A., & Ebrahimi, N.A. (2012). Lexical cohesion in English and Persian texts of novels. *Mediterranean Journal of Social Sciences*, 3. Retrieved January 3, 2014, from <http://www.mcser.org>.
- [26] Richards, J. C., Platt, J., & Platt, H. (1992). *Dictionary of Language Teaching and Applied Linguistics*. London: Longman.
- [27] Schwartz, M. (2006). Marks of Punctuation as False Grammatical Cognates. In B. J. Baer (Ed.), *Translating Russia: From Theory to Practice* (pp. 93–102). NY: Routledge.
- [28] Seddigh, F., Shokrpour, N., & Kafipour, R. (2010). Lexical cohesion in English and Persian abstracts. *Iranian Journal of Applied Language Studies*, 2. Retrieved January 6, 2014, from http://journals.usb.ac.ir/IJALS/en-us/Articles/Article_408.
- [29] Singh, R. (1979). *Contrastive Textual Cohesion*. Montreal: University de Montreal Publications.
- [30] Tarnyikova, J. (2009). *From Text to Texture*. Olomouc: FF UP.
- [31] Unesi, I. (1998). *Da'asta'an-e Dou Sha'hr* (Ch. Dickens, Trans.). Tehran: Negah Publications. (Original work published 1985)
- [32] Widdowson, H. G. (2007). *Discourse Analysis*. Oxford: OUP.
- [33] Wyrick, J. (2008). *Steps to Writing Well* (10th ed., Vol. 2). US: Thomson Wadsworth.
- [34] Xu, Y. C. (1996). A contrastive analysis of major cohesive devices in English and Chinese. *Shandong Foreign Language Teaching Journal*, 4, 1-6. Retrieved August 8, 2014, from <http://www.uri.edu/iaics>.
- [35] Yahaghi, M. J., & Naseh, M. M. (1996). *A Guide to Writing and Editing*. Mashhad: Astan Ghods Press.

Zahra Sojudifar was born in Bandar Abbass on 20 September 1984 and raised in Shiraz. She graduated from Bachelor Degree in English translation from Abadeh Azad University, Fars State, Iran on 2007 and then did an M.A on translation studies in Fars Research and Science Branch, Islamic Azad University, Fars State, Iran.

Azadeh Nemati is an Assistant Professor in Iran, majoring in ELT. She is the editor in chief of some international journals and has already published + 10 books and + 30 articles nationally and internationally. She has presented in many international conferences and also supervised some MA theses. In 2010, 2012 and 2014 she was selected as distinguished researcher in the University.

Mohammad Reza Falahati Qadimi Fumani is an Assistant Professor of Computational Linguistics. He currently works at the Department of Computational Linguistics at RICeST (www.ricest.ac.ir). He also teaches at different governmental and non-

governmental universities in Iran and acts as supervisor of M.A. theses on translation, linguistics and language teaching studies. He is also a researcher and has already published more than 30 articles, 10 books and has completed more than 10 research projects and has already held a number of workshop series for students of the aforementioned fields. His research hobbies are translation, linguistics, computational linguistics, language teaching and NLP (Natural language Processing).