Lexical Cohesion in English and Persian Texts of Novels

Ali Rahimi

Bahcesehir University (B A U), Istanbul, Turkey
Email: rahimijah@yahoo.com

Nabi A. Ebrahimi

Young Researchers Club, Arsanjan Branch, Islamic Azad University, Arsanjan, Iran
Corresponding Author: nabi.ebrahimi@iaua.ac.ir Tel: 00989171261337

Abstract: What are the stylistic differences between Persian and English novelists with regard to their choices and application of lexical patterns? What are the textual differences between Persian and English novels in terms of lexical cohesion? To answer these questions, two Persian (The Blind Owl by Sadigh Hedayat and The Patient Stone by Sadigh Chubak) together with two English novels (The Old Man and the Sea by Ernest Hemingway and The Pearl by John Steinbeck) which are enormously popular in contemporary Persian and English literatures were selected as the main source of data. The results of lexical cohesion analysis on 1000-word excerpts of these novels showed that the difference between Persian and English novelists' choices of lexical patterns is caused by the higher number of reiteration as well as the collocation pairs used by Persian novelists. This study concludes that some lexical relations are more frequent in the Persian novels than in the English ones (e.g. Equivalence and Elaborative collocation), conversely some are more salient in English texts under study (e.g. Simple repetition). The similarities of these texts with regard to the frequency of lexical relations turned out to be more than their differences nonetheless. In addition, detailed explanations of the results as well as complete elucidation of the model of analysis (proposed by Tanskanen, 2006) accompanied by examples extracted from the selected Persian and English novels have been put forward. The results of this study can inspire literary critics, theorists in stylistics, EFL teachers, assessment specialists, translators and contrastivists.

Keywords: stylistics, lexical cohesion, lexical relations, Persian novels, English novels

1. Introduction

Cohesion is the network of lexical, grammatical, and other relations which link various parts of a text. Halliday and Hasan (1976) take the view that the primary determinant of whether a set of sentences do or do not constitute a text depends on cohesive relationships within and between the sentences, which create texture. Cohesion as Halliday and Hasan define refers to relations of meaning that exist within the text and that define it as a text (Shafiee, 2003). The topic of cohesion has always appeared to be the most useful constituent of discourse analysis or text linguistics applicable to translation (Newmark, 1987). Analysis of cohesive devices within a text gives us more insight into how writers structure what they want to say and may be crucial factors in our judgments of whether something is well-written or not. Different languages may show a tendency towards the use of some specific types of different cohesive devices for creating texts. Such tendencies can be clarified by cohesion analysis. Among the different types of cohesive devices, i.e. reference, substitution, ellipsis, conjunction, and lexical cohesion (Haliday & Hasan, 1976), lexical cohesion is of great importance and has different applications in translation studies, computational linguistics, information retrieval researches and teaching and learning issues. Lexical cohesion is a type of cohesion whereby certain lexical features of the text connect sentences with each other in the text (Benbrahim, 1996, cited in Dahl, 2000, p.123). In this study, we focused on the lexical patterns in the texts of Persian and English novels. The present study is the first of its type and it is one of rare studies exploring stylistic properties of Persian novels. The rest of the article is presented as follows. At first, different applications of lexical cohesion analyses will be presented. Then different models for analyzing lexical cohesion along a full account of Tanksanen model for lexical cohesion analyses (used in the study) are introduced. The results of the study and a full discussion of the findings form the following parts. Concluding remarks are presented in the last part.

Applications of lexical cohesion analyses

Lexical cohesion analyses can be of different applications including summarization, segmentation, anaphora resolution, improved speech recognition, improved optical character recognition and improved machine translation.

A rudimentary summary of a document can be created by extracting the most important sentences from the document, where the importance of a sentence is measured by the presence of key phrases. Such summaries often contain outliers, sentences that do not fit with the other sentences. A measure of lexical cohesion can be used to detect and remove these outliers, thereby improving the quality of the summary.

According to Berber (1999), segmentation refers to the principled division of texts into contiguous constituents. He added that segmentation means using the computer to divide a written text into acceptable parts (or segments). Among the different kinds of segments, lexical segments are of great importance.

A typical document first introduces an entity, such as a company, by giving its full name. Later in the document, the entity will be mentioned more briefly, using phrases like "the company" or simply "it". Anaphora resolution is the task of recognizing that these shorter phrases refer to the same entity as the full name. One approach to anaphora resolution involves building a chain of lexically cohesive terms, connecting sentences in the given document that discuss the same entity.

A measure of lexical cohesion can be used to recognize when speech recognition software has made errors. The incorrect words usually do not cohere with the rest of the text. Errors in optical character recognition can also be detected by their lack of lexical cohesion. Errors in machine translation also lack cohesion, although they may be more cohesive than errors in speech recognition and optical character recognition.

Among these applications, summarization and also segmentation have been of great interest for researchers in the field of corpus linguistics and language study. Various researchers (e.g. Benbrahim (1996, cited in Dahl, 2000, p.123) and Ercan & Cicekli (2008) have considered summarizing text by the use of lexical cohesion. Benbrahim (1996, cited in Dahl, 2000, p.123) stated that an analysis of lexical cohesion in text, primarily by counting repetitions, synonyms and paraphrase, leads to the establishment of a network of sentences, some tightly bonded through lexical cohesion relations, some others having weak bonds or no bonds at all. The strength of connections in this cohesion network is used to determine key sentences in a text.

Lexical cohesion is also of great importance in translation studies and pedagogical issues. Within a discoursal framework, Lotfipour-Saedi (1997) introduced seven dimensions of translation equivalence (TE): vocabulary, structure, texture, degree of indirection, language variety, cognitive effect and aesthetic effect. He listed four components for texture: thematization strategies, schematic structures, cohesion, and paralanguage. Then he talked about lexical cohesion and its implication for TE. He added that in evaluating a translated text to determine the degree to which the translator has managed to maintain the discoursal value intended by the SL text, especially in literature-texts, lexical cohesion can offer critical guidelines.

Lexical cohesion is also of great importance for language teachers, material developers and students. McCarty (1991) stated that it is necessary to give a discourse dimension to vocabulary teaching and vocabulary activities in the classroom. He continued that Halliday and Hasan's (1976) description of lexical cohesion is one recent attempt at studying vocabulary patterns above sentence level. He added that the study of vocabulary in discourse is concerned with patterns in text created by the vocabulary or lexical relations that are found over clause and sentence boundaries, the role of special words in organizing discourse and signaling their structure and the relationship between these features of textuality and the register of the end product. Students themselves can be motivated to collect items along discourse-functional lines, something which become more and more important as they embark on composition writing and argumentation in general. In addition, contrastive analysis of different discourse features of two languages can be inspiring for the teachers and material developers to find the difficulties the learners may face through language learning. Studies such as Salehuddin et.al, (2006) and Forutan & Nasiri (2011) tried to spot the difficulties may face in their transfer of L1 features to L2.

The model of analysis used in this study

The model utilized in this study was developed by Tanskanen (2006). She borrowed some of the categories from the previous models (e.g. Halliday & Hasan, 1976 and Morris and Hirst (1991)) and believed that her model "provide a good basis for understanding the work done by lexical cohesion in discourse" (Tanskanen, 2006, p.49).

Let us now turn to the categories of lexical cohesion proposed by Tanskanen (2006) and discuss them in detail with examples. The categories are:

Reiteration:
simple repetition
complex repetition
substitution
equivalence
generalization
specification
co-specification
contrast
Collocation:
ordered set
activity-related collocation
elaborative collocation

Reiteration relations
Simple and complex repetition

Here repetition is divided into simple repetition and complex repetition. Tanskanen (2006) asserted that simple repetition occurs when an item is repeated either in an identical form or with no other than a simple grammatical change, e.g. singular – plural, present tense – past tense. Complex repetition involves a more substantial change: the items may be identical but serve different grammatical functions, or they may not be identical but share a lexical morpheme.
The following example selected from The Pearl includes both simple and complex repetition:
….Kino *saw* Juana arise silently from beside *him. He saw her move* towards the fireplace. So carefully did *she* work that *he* heard only the slightest sound when *she moved* the fireplace stone.
In this example, we find both simple and complex repetition at work. *Saw-saw, him-he, her-she, he-he and she-she* are instances of simple repetition, while *move-moved* is an example of complex repetition. So in this example, there are five pairs of simple repetition and there is one pair of complex repetition. Studying lexical cohesion in multiparty conversations in English, Gonzalez (2011) reports repetition (59%) as the most frequent lexical cohesion device.

Substitution

The second category of reiteration relation is substitution. The most usual form of substitution is a pronoun substituting for a noun. In the following sentences selected from The Old Man and the Sea, *they* reiterates *my eyes* or *they* substitutes for *my eyes.*

All my life the early sun has hurt my eyes, he thought. Yet they are still good.
In the next sentences selected from the Persian novel "the Patient Stone", *eu* substitutes for *Gohar*
Gohar kolfate shomas. Eu kasio nadare

Behjat (2009) studied cohesive devices in English novels and their corresponding Persian translations and mentioned that pronouns are used in English more than Persian.
Beside pronouns, there are other substitutions items as well, like *one, do* and *so*, which can substitute for previous items. These kinds of substitution are actually relatively rare in the material.

Equivalence

The third subcategory of reiteration in our classification is that of equivalence. Following McCarthy (1988), the term equivalence is used to refer to the relation more commonly referred to as synonymy. I should mention again that our approach is a discourse-specific one to lexical relations. As Tanskanen (2006) stated that the significance issue for this more discourse- specific approach to lexical relations is the language user's decision to use an item for instance in equivalence with another item, although they may not be semantically absolutely synonymous. Using non-lexical-semantic terms draws attention to the fact that the justification and explanation for a relation between lexical items can and should be sought for in the text in which the items occur.
Let us consider some examples. In the following sentences selected from the Persian novel "the Blind Owl", *cheke cheke* and *ghatre ghatre* are in an equivalence relation.

Hala mikhaham sartasare zendegiye khodam ra manande khusheye angur dar dastam befesharam va osareye anra- na, sharable anra- *ghatre ghatre* dar galuye khoshke sayeam mesle abe torbat bechekanam........mikhaham osareye- na, sharable talkh-e- zendegiye khodam ra *cheke cheke* dar galuye khoshke sayeam chekanide be eu beguyam: "in zendegiye man ast!"

Rescue and save in the next sentences selected from The Pearl are in an equivalence relation.

All of the time Juana had been trying to *rescue* something of the old peace, of the time before the pearl. But now it was gone, and there was no retrieving it. And knowing this, she abandoned the past instantly. There was nothing to do but to *save* themselves.

Generalization

The fourth subcategory is *generalization,* which covers the relation between an item and a more general item. This relation has been referred to as a *superordinate* or *hyponymic* relation in most of the earlier studies, or *inclusion: specific-general* by McCarthy (Tanskanen, 2006).

In the following sentence selected from "the Patient Stone", *bahr* reiterates and generalizes *gohar*

Tora setayesh mikonam ke yekta gohar-e bahr-e mavaje afarinesh maeim

In the following text, political party is a generalization of labor:

Gordon: If *Labor* get in and they can't fulfill their promises...

 (21 turns)

Audrey: Well I can't well I mean there's an awful lot, I mean would, no no matter which *political party* it is, they all make promises, but they don't carry them all out.

Specification

Specification, the fifth subcategory, is the opposite of generalization: it refers to the relation between an item and a more specific item. This relation has previously been called *meronymy,* and McCarthy referred to it as *inclusion: general-specific* (Tanskanen, 2006).

In the following sentences, *health, education* reiterate and specify *the other social services.*

The deceptive nature of the accelerated growth argument occurs also with respect to *the other social services.* The White Paper tells us that what we want to do in *health, education,* etc. depends on faster growth.

Haj esmoeil is a specification of *bozorguna* in the next sentences selected from the Patient Stone.

Man ye dokhtare rakhtshuri budam ke nanam esmesh khojaste bud va too khunehaye *bozorguna* rakhtshuri mikard. Yeki az unaee ke nanam miraft khunashun rakhshuri mikard hamin *haj esmoeil* tajere gomrok bud ke tu gomrok hojre dasht.

Co-specification

The next subcategory of reiteration is *co-specification,* which includes the relation between two items which have a common general item. The earlier studies that have included this relation have referred to it as *co-hyponymy* or *co-meronymy.*

The general item would be *tan* for the items *khun*, *sar*, *dasthayam* in the next sentences selected from The Blind Owl:

Vali man bidar budam, hes mikardam tanam dagh hast va lakehaye *khun* be aba va shal gardanam chasbide bud. *Dasthayam* khunin bud. Ama ba vojude tab va davare *sar*, yekno ezterab va hayajane makhsus dar man tavalod shode bud ke shadidtar az fekre mahv kardane asare *khun* bud......

In the next example, *RP speakers* and *Standard English speakers* are related even without the mention of the general item *English speakers.*

It is widely agreed, though, that while all *RP speakers* also speak Standard English, the reverse is not the case. Perhaps 9%-12% of the population of Britain (see Trudgill & Cheshire 1989) speaks Standard English with some form of regional accent. It is true that in most cases *Standard English speakers* do not have 'broad' local accents (i.e. accents with large numbers of regional features which are phonologically and phonetically very distant from RP)...

Contrast

The final subcategory of reiteration is *contrast,* which refers to the relation between an item and another item which has an opposite meaning. This relation has also been called *antonymy*, *opposition,* or *complex repetition*

or paraphrase.

In the following sentences selected from The Blind Owl, *javan* and *pirmard* are related by contrast.

Har kas diruz mara dide bud, *javan*-e shekaste va nakhoshi did ast. Vali emruz *pirmard*-e ghuzi mibinad ke muhaye sefid, cheshmhaye vasukhte va labe shekari darad.

In the following example borrowed from Tanskanen (2006), the contrastiveness of *out of fashion* and *up to date is* enhanced by the use of, respectively, *dramatically* and *completely.*

And the reason for this is that it belongs to a tradition, a fashion if you like, of writing which went dramatically out of *out of fashion* immediately after World War One. So, at the time when it was published most readers would have regarded it as completely *up to date* in its in its style and in its presentation.

Collocation relations

Now it is time to define and discuss collocation relations. This category has been notoriously to define, so much so that it has often been omitted from analysis.   . Regardless of the difficulties that we will no doubt have to face, collocation relations will be included in our model of analysis. The three types of collocations are presented in the following sections.

Ordered set

Ordered set is the first of the collocation subcategories in our classification. It is perhaps the clearest of the three categories, the easiest to detect and closest to the more systematic reiteration relations described above. The category includes members of ordered sets of lexical items, for example, colors, numbers, months, days of the week and the like. Because the sets are relatively clear, these relations are not difficult to find in texts, but as luck would have it, they seem to be quite infrequent: there are only a few instances of ordered set in the material of the present study (Tanskanen, 2006).

In the following sentences selected from The Blind Owl, we find *ruz, saat, mah, sal* as ordered set.

Shayad az anjaee ke hameye ravabete man ba donyaye zendeha boride shode, yadegarhaye gozashte jeloam naghsh mibandad! Gozashte, ayande, *saat, ruz, mah* va *sal* hame barayam yeksan ast.

Second, third and *fourth* in the next sentences from The Old Man and The Sea are considered as ordered set.

One bait was down forty fathoms. The *second* was at seventy five and the *third* and *fourth* were down in the blue water one hundred and one hundred and twenty-five fathoms.

Activity-related collocation

The study of Martin (1992), which presents a redefinition of Halliday and Hasan's collocation category, divides these relations into two: *nuclear (extending* and *enhancing)* and *activity sequence* relations. Let us therefore concentrate on what Martin (1992) calls nuclear relations, which reflect the ways in which "actions, people, places, things and qualities configure as activities". As an example of such relations Martin mentions, among others, *serve — ace.*

It appears that in the present material as well we can find pairs such as *cyphers — decode* or *meals—or driving— the same car* in which the relation between the items is based on an activity: you can *decode* cyphers, eat meals and drive cars. In classifying such items, it may thus be helpful to think of their association as resulting from such a relation. The category will consequently be called *activity-related collocation*

mive and *nachinish* in the next sentences from the Patient Stone are in such a relation.

Dokhtare balegh mesle *mive*-ye reside immune, age *nachinish* khodesh legh mishe miofte va dige bedarde sahib nemikhore.

The following example contains such a relation between *cyphers — decode.*

C: well I expect you don't need *cyphers* during if by that you mean people who

e: people who can *decode* yeah

Elaborative collocation

There are other items between which an association exists but which cannot be grouped as ordered set or activity-related collocation. Consider the following example quoted from Tanskanen (2006).

... at the beginning of the Michaelmas term 1955, Sylvia's first year at *Cambridge. I* had walked into *the Mill Lane lecture*

room a few minutes early...

The relation in the above example illustrates our third collocation category, *elaborative category*. This is a category for all those pairs whose relation is not possible to define more specifically than stating that the items can somehow be developed on the same topic.

You should keep in mind that the relation between items is not completely haphazard. It is with elaborative collocation that the frame concept is introduced. Frames are knowledge structures evoked by lexical items: for example, if a text begins with *arraignment,* it evokes the arraignment frame, and following items, such as *magistrate* and *charges* are interpreted according to this frame, thus creating coherence in the text (Fillmore & Baker 2001). Considering the above example, we can say that *Cambridge* evokes the university frame, and *the Mill Lane lecture room* can be interpreted within this frame.

Tanskanen (2006) said that a "trigger-test" can be helpful in verifying some elaborative relations. Trigger is a concept introduced by Jordan (1998). Jordan argued that a trigger, which is usually a repetition of the previous topic (item), can be used to clarify the association between an item and its re-entry. Let us consider our example in a slightly modified form:

... at the beginning of the Michaelmas term 1955, Sylvia's first year at *Cambridge. I* had walked into *Cambridge's Mill Lane lecture room* a few minutes early...

If there was no relation between the items, the result with the trigger would hardly be so satisfying. So, "trigger-test" is of great importance for detecting this kind of collocation.

Let us consider two examples which are selected from our representative Persian and English novels.

Nevisandegi che honare naghesiye.

This sentence has been selected from the Patient Stone. The relation between *nevisandegi* and *honar* can be considered as elaborative collocation due to the satisfying result of the trigger-test: honare nevisandegi

The pale *moon* dipped in and out of strands of clouds so that Juana walked in darkness for a moment and in *light* the next.

In the above sentence taken from The Pearl, *moon* and *light* are in elaborative collocation relation.

2. Method

Two Persian and two English novels were selected as the main sources of data for this study. The selected Persian novels were The Blind Owl by Sadigh Hedayat and The Patient Stone by Sadigh Chubak which are among the most famous novels in modern Persian literature. The Old Man and the Sea by Ernest Hemingway and The Pearl by John Steinbeck were selected as representative novels which are of great reputation in contemporary English literature. Then a 1000-word excerpt was extracted from each novel. The selection of these four excerpts was completely random. Finally, types of lexical cohesion relations in each excerpt were detected according to the model of lexical analysis proposed by Tanskanen (2006) and their numbers were recorded.

3. Results and discussion

In the present study, the number of lexical cohesive ties was manually detected based on the proposed model through the selected texts. This section concentrates on the variation found in the use of lexical cohesion relations in the selected texts of Persian and English novels and on the potential differences between reiteration and collocation. Quantitative information on the occurrences of pairs of reiteration and collocation in the novels is tabulated in table 1 to 4.

Table 1. Pairs of reiteration and collocation (per 1000 words) in the Persian and English novels

	Persian novels(mean)	English novels (mean)
Reiteration	268	250
Collocation	22	14.5
Total	288	264.5

Table 2. Mean of pairs of reiteratio relations and collocation (per 1000 words) in the two selected Persian novels: The Blind Owl and The Patient Stone.

	Persian novels (mean)	English novels (mean)
Simple repetition	173	186
Complex repetition	14	8
Substitution	38	36
Equivalence	12	2
Generalization	6	1.5
Specification	13	8.5
Co-specification	5	2
Contrast	7	6
Ordered-set	2	1
Activity-related collocation	5.5	5
Elaborative collocation	14.5	8.5

Table 3. Pairs of reiteration and collocation relations (per 1000 words) in the two selected groups of Persian and English novels

	The Blind Owl	The Patient Stone	Mean
Simple repetition	182	164	173
Complex repetition	8	20	14
Substitution	25	51	38
Equivalence	19	5	12
Generalization	5	7	6
Specification	22	4	13
Co-specification	6	4	5
Contrast	7	7	7
Ordered-set	3	1	2
Activity-related collocation	6	5	5.5
Elaborative collocation	12	17	14.5

The numbers shown in the tables are normalized frequencies of the cohesive pairs occurring in the Persian and English novel texts. In other words, frequencies of cohesive pairs are normalized to a text of 1000 words extracted from each novel, showing how many times they would occur in a text of 1000 words. This makes it possible to compare frequencies of cohesive pairs in the selected texts in a reliable way. The numbers in the tables are therefore directly comparable, which facilitates comparing our findings.

Table 3 shows the number of different lexical relations in the 1000-word excerpts extracted from our two Persian novels: The Blind Owl and The Patient Stone. Table 4 is related to the 1000-word excerpts extracted from the English novels: The Old Man and the Sea, and The Pearl.

Let us start with the distributions of pairs of reiteration and collocation in the selected novels. As you can see, reiteration relations are more frequent both in Persian and English novels. There are 268 reiteration pairs (i.e., simple repetition, complex repetition, substitution, generalization, specification, equivalence, co-specification, contrast) in 1000-

word texts selected from Persian novels in average and 250 in selected texts of English novels. The average number of collocation pairs (Ordered-set, elaborative collocation, activity-related collocation) is 22 and 14.5, respectively in Persian and English novel texts.

Table 4. Pairs of reiteration and collocation relations (per 1000 words) in the two selected English novels: The Old Man and the Sea and The Pearl

	The Old Man and the Sea	The Pearl	Mean
Simple repetition	199	173	186
Complex repetition	6	10	8
Substitution	24	48	36
Equivalence	2	2	2
Generalization	2	1	1.5
Specification	9	8	8.5
Co-specification	3	1	2
Contrast	6	6	6
Ordered-set	0	2	1
Activity-related collocation	7	3	5
Elaborative collocation	7	9	8.5

The two novel groups thus seem to differ as regard to the total number of pairs; the difference is caused by the higher number of reiteration and also collocation pairs in Persian novel texts.

We can then move on to examine the distribution of cohesive pairs in more detail, i.e. per categories of reiteration and collocation.

The most frequent relations in both of the Persian and English novel groups are simple repetition relations. Substitution comes second in both Persian and English novel texts. The order of frequency with regard to the rest of the relations for the Persian novel texts is elaborative collocation, complex repetition, specification, equivalence, contrast, generalization, activity-related collocation, co-specification and ordered-set. The same order for English novel texts is elaborative collocation, specification, complex repetition, contrast, activity-related collocation, co-specification, equivalence, generalization and ordered-set.

Persian and English novel texts are closely similar with regard to the average number of simple repetition, substitution, contrast, ordered-set and activity-related collocation relation pairs. The most striking difference between these novels with regard to the average number of cohesive pairs is related to equivalence. The frequency of pairs in equivalence relation is much higher in Persian novel texts. The number of pairs having generalization relation is also much more in Persian novel texts. Specification relations are also less frequent in English novel texts compared to Persian novel texts. Complex repetition and co-specification relations are also more frequent in Persian novel texts compared to the similar texts in English.

4. Conclusion

Cook (1989) believes one should know about cohesive ties and their stylistic role in discourse. Among the different types of cohesive devices, i.e. reference, substitution, ellipsis, conjunction, and lexical cohesion (Haliday and Hasan, 1976), lexical cohesion is of great importance and has different applications in translation studies, computational linguistics, information retrieval researches and teaching and learning issues. The cognitive structure of the reader of a text can be manipulated by various lexical cohesion strategies. It is the responsibility of the translator to be aware of these strategies and to reflect them in their translation. Creating summaries and also segmenting texts into thematically coherent units are mainly performed by the use of lexical cohesion. Lexical cohesion is also of great importance for language teachers, material developers and students.

In this study, we focused on the lexical patterns in the texts of Persian and English novels. Two Persian and two English novels were selected as the main sources of data for this study. The selected Persian novels were The Blind Owl

by Sadigh Hedayat and The Patient Stone by Sadigh Chubak. The Old Man and the Sea by Ernest Hemingway and The Pearl by John Steinbeck were selected as representative novels which are of great reputation in contemporary English literature. Then a 1000-word excerpt was extracted from each novel. The number of lexical cohesive ties was manually detected based on the model proposed by Tanskanen (2006) through the selected texts. The focus was on the variation found in the use of lexical cohesion relations in the selected texts of Persian and English novels and on the potential differences between reiteration and collocation. Quantitative information on the occurrences of pairs of reiteration and collocation in the novels was provided.

The result of this study showed that the two novel groups seem to differ as regard to the total number of pairs; the difference is caused by the higher number of reiteration and also collocation pairs in Persian novel texts. The most frequent relations in both of the Persian and English novel groups are simple repetition relations. Substitution comes second in both Persian and English novel texts. The order of frequency with regard to the rest of the relations for the Persian novel texts is elaborative collocation, complex repetition, specification, equivalence, contrast, generalization, activity-related collocation, co-specification and ordered-set. The same order for English novel texts is elaborative collocation, specification, complex repetition, contrast, activity-related collocation, co-specification, equivalence, generalization and ordered-set.

Persian and English novel texts are closely similar with regard to the average number of simple repetition, substitution, contrast, ordered-set and activity-related collocation relation pairs. The most striking difference between these novels with regard to the average number of cohesive pairs is related to equivalence. The frequency of pairs in equivalence relation is much higher in Persian novel texts. The number of pairs having generalization relation is also much more in Persian novel texts. Specification relations are also less frequent in English novel texts compared to Persian novel texts. Complex repetition and co-specification relations are also more frequent in Persian novel texts compared to the similar texts in English.

References

Behjat, F. (2009). Cohesive devices in English novels and their corresponding Persian translations. *Journal of Teaching English as a Foreign Language and Literature, 1*(3), 107-118.

Berber Sardinha, T. (1999). Looking at discourse in a corpus: The role of lexical cohesion. *Proceedings of AILA 2000 on CD-ROM, AILA - 12th World Congress of Applied Linguistics* (pp. 7-13). Tokyo.

Cook, G. (1989). *Discourse.* Oxford: Oxford University Press

Dahl, T. (2000). Text Summarisation: From Human Activity to Computer Program. The Problem of Tacit Knowledge. *Hermes, Journal of Linguistics, 25*, 113-131.

Ercan, G. & Cicekli, I. (2008). Lexical cohesion based topic modeling for summarization. In *Proceedings of the Cicling 2008*, pp. 582–592.

Fillmore, C.J. & Baker, C.F. (2001). Frames and the semantics of understanding. Retrieved 12 October 2008, from http://framenet.icsi.berkeley.edu/

Forutan, A. & Nasiri, S.R. (2011). Signaling Nouns in English and Persian: A Contrastive Study. *Theory and Practice in Language Studies, 1*(10), 1273-1283

Gonzalez, M. (2011). Lexical cohesion in multiparty conversations. *Language Sciences, 33*(1), 167-179.

Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English.* London: Longman

Jordan, M.P. (1998). Pragmatic, stylistic and grammatical limitations on choice: a study of cause-effect signaling in English. In A.Sanchez-Macarro & R.Carter (Eds), *Linguistic Choice Across Genres: Variation in Spoken and Written English* (pp.65-86). Amsterdam & Philadelphia: John Benjamins.

Lotfipour-Saedie, K. (1997). Lexical Cohesion and Translation Equivalence. *Meta, XLII* (I), 185-192.

Martin, J.R. (1992). *English Text: System and Structure.* Amsterdam & Philadelphia: John Benjamins.

McCarty, M. (1991). *Discourse Analysis for Language Teachers.* Cambridge University Press.

Morris, J. & Hirst, G. (1991). "Lexical cohesion computed by thesaural relations as an indicator of the structure of text". *Computational Linguistics, 17*, 21-48.

Newmark, P. (1987). The use of systemic linguistics in translation analysis and criticism. In Steele, R. & T. Thtreadgold (Eds), *Language Topics: Essays in Honour of Michael Halliday, 1*, (pp. 293-303), Amsterdam/Philadelphia: John Benjamins.

Salehuddin, K., Tan Kim, H. & Marlyna M. (2006). Definiteness and indefiniteness: a contrastive analysis of the use of determiners between the Malay language and English. *GEMA Online Journal of Language Studies, 6*(1), 21-30.

Shafiee, N.S. (2003). *Grammatical Cohesion in English and Persian in Contrast.* Retrieved 23 October 2008, from http://www.kkhec.ac.ir/klf/shafiee.htm.

Tanskanen, S.K. (2006). *Collaborating towards Coherence.* Amsterdam & Philadelphia: John Benjamins.