Lexical Cohesion in English and Persian Abstracts

Fatemeh Seddigh

Shiraz University of Medical Sciences, Iran

fseddigh.english@yahoo.com

Nasrin Shokrpour

Shiraz University of Medical Sciences, Iran

shokrpourn@gmail.com

Reza Kafipour

Islamic Azad University, Shiraz Branch, Iran

rezakafipour@gmail.com

Abstract

This study compares and contrasts lexical cohesion in English and Persian abstracts of Iranian medical students' theses to appreciate textualization processes in the two languages. For this purpose, one hundred English and Persian abstracts were selected randomly and analyzed based on Seddigh and Yarmohamadi's (1996) lexical cohesion framework, a version of Halliday and Hasan's (1976) and Halliday's (1985) taxonomies. For contrastive analysis, the SPSS package was used. The results revealed some similarities and differences in the use of lexical cohesion sub-categories in the parallel English and Persian texts. The occurrence of all sub-types is nearly the same in the two groups of texts and the two- tailed t-test employed showed that the differences are not statistically significant. Both languages exhibit a general tendency in using repetition, but synonymy and meronymy are the least used sub-categories. Regarding the density of the texts, the analysis indicates that Persian abstracts are denser than their corresponding English ones.

Introduction

Cohesion is the semantic relation that exists within a text and refers to the way sentences are connected to each other to produce a text. According to Morris and Hirst (1991: 10), "cohesion is the textual quality responsible for making the sentences of a text seem to hang together".

Cohesive devices or "ties" are words, phrases or clauses that link the items together, and cohesive relations are the ways two or more items are semantically joined to each other in a text. These relations might be grammatical or lexical (see Halliday, 1989: 49). Although lexical cohesion is more pervasive in creating textual cohesion, it is somehow neglected in EFL as noted by Flowerdew (2006: 209). Various cohesive devices are used in different languages to achieve textuality; however, some languages may show a tendency towards the use of some specific kinds of them. These tendencies can be illuminated by cohesion analysis and this study is devoted to lexical contrastive cohesion analysis. Lexical cohesion contributes to cohesion by the selection of vocabulary.

According to Morris and Hirst (2003: 1), "lexical cohesion occurs when related word pairs join together to form larger groups of related words that can extend freely over sentence boundaries". These lexical semantic relations or "term relationships" assist in providing the continuity of lexical meaning in a text. They are created by repetition or by using semantically related words in a text. The relations may consist of synonymy or near synonymy, e.g. sick and ill; hyponymy or specific-general relationship, e.g. pediatrician and doctor; meronymy or part-whole relationship, e.g. leg and body; antonymy or opposite words, e.g. malignant and benign; collocation which is the high tendency of lexical items to co-occur, e.g. patient and disease and general nouns which have generalized references, e.g. human.

Different types of lexical cohesion play an important role in the textualization of texts (see Francis 1994), and various languages employ them differently. While lexical contrastive cohesion analysis is not a new subject, its application to English and Persian has been minimal (e.g. Ikegami, 1980; Johns, 1980; Evans et al., 1983; Linnarud, 1987; Mohebbi, 1998; Nourmohammadi, 1988; Ziahosseiny, 1999).

Sahragard (1992) selected his data from among contemporary Persian and English plays as well as Persian translations of English plays and studied lexical cohesion in them. The results show that in all the three sets of texts, repetition is the most frequent sub-type. Tanskanen (1995) compared the frequency and the density of lexical cohesion in three different genres of conversation, academic prose and prepared speech. He notices that reiteration is more frequent in all genres. Regarding the density of lexical cohesion, prepared speech is the most dense and conversation is the least dense genre. He also concludes that reiteration density increases as conditions become more difficult. Seddigh and Yarmohammadi (1996) performed a contrastive study on lexical cohesion in parallel English and Persian journalistic political texts. The results indicate that although both groups of texts manifest a general tendency towards the use of repetition, Persian texts are denser in this regard. The occurrence of other sub-types are nearly the same in both groups of texts and the densities of parallel English and Persian texts regarding the use of lexical ties are almost the same. Taboada (2000) studied cohesion in Spanish and English task-oriented dialogues. One of her conclusions is that lexical cohesion, and specifically repetition of the same item is the most widely used type. Veisi (2002) carried out a lexical contrastive cohesion analysis in literary English texts and their corresponding Kurdish translations. The results indicate that both groups of texts show a general tendency toward utilizing repetition but

Kurdish texts are denser in this regard. The difference between the density of lexical cohesion in English and Kurdish texts is not statistically significant.

Mirshekaran (2003) studied lexical cohesion in Persian narrative and expository texts types. The results show that the frequency of occurrence of repetition is more in expository texts than in narrative ones.

Regarding the density of the texts, expository texts are tighter than the narrative ones. Castro (2004) carried out a study to find the relation between cohesion and the social construction of meaning in the essays written by Filipino college freshmen. One of the results was that lexical repetition and use of synonyms were the commonest device for creating cohesion. MacMillan (2007) found that lexical cohesion has a basic role in reading comprehension sections of different types (paper-based, computer-based or internet-based) of TOEFL test. The results show that all reading comprehension questions on the test involve the identification of different cases of lexical repetition. Kai (2008) compared and contrasted the dissertation abstracts of native and non-native speakers of English in the field of applied linguistics. He noticed that while native speakers employed more complex repetition in their abstracts than non-native speakers, the use of simple repetition had a higher rate in non-native speakers' abstracts.

The present study aims at determining the frequency and the degree of utilization of lexical cohesion sub-types in English and Persian on the basis of corresponding English and Persian abstracts of theses written by Iranian medical students. Moreover, it compares the density of the texts in both languages regarding the use of lexical ties. The results of this study can be used as a basis for a theoretical description of lexical cohesion in Persian. Moreover, they can be utilized for pedagogical and translation purposes.

Material and Method

The data consist of English and Persian abstracts of medical students' theses.

One hundred English and Persian abstracts were selected randomly from 2000 to 2005 and analyzed. In each case the first 250 words of the Persian abstracts and their corresponding English ones were subjected to analysis. The total number of words analyzed amounted to about 25000 in number, 12500 in each language roughly.

The data were analyzed based on Seddigh and Yarmohammadi's (1996) framework which is a version of Halliday and Hasan's (1976) and Halliday's (1985) lexical cohesion taxonomies. Halliday and Hasan (1976) consider reiteration and collocation as two main categories of lexical cohesion.

Reiteration is divided into the same word, synonymy, superordinate and general words. Halliday's (1985) taxonomy consists of repetition, synonymy and collocation as three main categories. Synonymy is divided into hyponymy, meronymy and antonymy. In Seddigh and Yarmohammadi's (1996) framework, repetition, synonymy, hyponymy, meronymy, antonymy and general nouns are considered as parts of reiteration. The second major category is collocation which is achieved through the association of lexical items that regularly co-occur. The following examples are provided for illustration:

Repetition (R)
ex. Malabsorption syndroms are mostly <u>diagnosed</u> in children. In most cases, duodenal biopsy helps the <u>diagnosis.</u>

Synonymy or Near–Synonymy (S)
ex. Due to the importance of neck <u>traumas</u>, ----------------------.
Most <u>injuries</u> were to the left of the neck II area.

Hyponymy (Specific- General) (H)

ex. Head and neck squamous cell <u>carcinoma</u> is noticed in ------------.

It is the most common <u>cancer</u> that affects ------------------.

Meroneymy (Part-Whole) (M)

ex. Thalassemia is the most common inherited blood disorder around the <u>world.</u> Our country, <u>Iran,</u> is located in the endemic region of B- thalassemia.

Antonymy (Opposites) (A)

ex. The <u>severe</u> form of the disease is caused by ---------------. The <u>mild</u> cases are characterized by --------------.

General Nouns (General words as cohesive elements) (G)

ex. <u>Hepatitis C virus infection</u> is common among --------.

The <u>condition</u> is characterized by ----------.

Collocation (Co-occurrence tendency) (C)

ex. The <u>risk factors</u> of the disease may be internal or external.

For contrastive analysis of the obtained data, the SPSS package was employed. For each sentence, the number of lexical ties, types of lexical cohesion, and the distance between presupposed and presupposing items were exploited. In order to determine the frequency and degree of utilization of various types of lexical cohesion in each group of texts, descriptive statistics was used, i.e. the percentages of different types of lexical cohesion in each language were computed. Also, a two-tailed t-test was carried out to determine whether the difference between the mean percentages of the sub-types in the two groups of abstracts is statistically significant or not. Furthermore, the densities of the two groups of texts regarding the use of lexical ties were given the same treatment.

Analysis and Results

Table 1 shows the frequency of lexical cohesion sub-types in English and Persian texts.

Type of lexical cohesion	R	S	M	C	H	G	A
Persian texts	63.42	2.29	2.49	7.36	4.19	15.23	5.03
English texts	63.76	2.61	2.58	6.76	4.41	14.1	5.77
T $_{obs}$ values	-0.38	-0.81	-0.39	1.02	-0.62	1.45	-1.01

Table1. The average percentages of lexical cohesion sub-types in English and Persian

$P < .05$, two- tailed

R= repetition   S= synonymy   M= meronymy   C= collocation

H= hyponymy   G= general noun   A= antonymy

As displayed in Table 1 and diagrams 1 and 2 (see the Appendix), in Persian texts repetition is the most frequently used sub-type of lexical cohesion.

The next frequent sub-type is general noun, followed by collocation, antonymy, hyponymy, meronymy and synonymy in decreasing percentages of occurrence, i.e. R, G, C, A, H, M, S. In the case of English texts, repetition has the highest percentage of occurrence as well. The next frequently used sub-type is general noun, followed by collocation, antonymy, hyponymy, synonymy and meronymy, i.e. R, G, C, A, H, S, M. It is noteworthy that the orders of the sub-types are almost the same in both languages, except in the case of M and S. It can also be noticed that the densities of the sub-types vary within each group of texts.

The results indicate that the occurrences of all sub-types are almost the same in both languages. The two groups of texts exhibit a general tendency towards the use of repetition, i.e. 63.42% vs. 63.76%, but synonymy and meronymy play minor roles in producing cohesion in both languages (about 2%).

To see whether the differences between the mean percentages of lexical cohesion sub-types in English and Persian abstracts are statistically significant or not, for each case a two-tailed t-test was run and the observed values of t were computed and compared with the t critical value at .05 level of significance. For this level, the critical value, with regard to the number of degrees of freedom, is 2.021 in the t-distribution Table. As the figures in Table 1 show, in all cases the differences are not statistically significant. So, it can be concluded that the occurrences of all sub-types are approximately the same in both groups of texts.

The densities of lexical cohesion in the texts were determined by dividing the total number of lexical ties in each language by the total number of sentences in that language. The obtained figures are 6.59 and 5.07 for the Persian and English texts, respectively. By doing a two- tailed t-test, it became clear that the difference is statistically significant. Thus, Persian texts are denser than their corresponding English ones.

Conclusion

The contrastive study of lexical cohesion in Persian and English abstracts of Iranian medical students' theses reveals that the occurrence of all the sub-types of lexical cohesion devices as well as their orders are almost the same in both languages. In Persian texts, R, G, C, A, H, M, S and in English ones R, G, C, A, H, S, M appear with decreasing percentages of occurrence, respectively. In both groups of texts, repetition is the most but synonymy and meronymy are the least frequently used sub- types. And finally, the densities of the texts regarding the use of lexical ties are not the same. The application of a two-tailed t-test revealed that the difference was statistically significant and Persian texts are denser than their corresponding English ones.

Pedagogical Implications

The results of this study reveal that Iranian medical students tend to employ a considerable amount of repetition in their abstracts without any variety. In other words, their writing lacks complexity of vocabulary which is very important in producing fluency in writing. Therefore, they should be thought to use more morphologically complex words, i.e. complex repetition instead of a simple word. Writing exercises can be designed to teach cohesion and exercises which demand the learners form different sub-types of lexical cohesion devices such as hyponymy, general nouns, autonomy, etc. In general, clause and sentence structure exercises should not ignore the textual consideration and more emphasis should be put on lexical cohesion through carefully constructed exercises.

References

Castro, C. (2004). "Cohesion and the social construction of meaning in the essays of Filipino college students writing in L2 English", *Asia Pacific Education Review*, 5(2), pp. 215-225.

Evans, M., Markowitz, J., Smith, R., Warmer, O. (1983). *Lexical Semantic Relations: A Comparative Survey*, Edmonton, Alberta: Linguistic Research Inc.

Flowerdew, J. (2006). "Use of signaling noun in a learner corpus. Lexical Cohesion and Corpus Linguistics (Special Issue)", *International Journal of Corpus Linguistics,* 11(3), pp. 227-247.

Francis, G. (1994). Labelling discourse: An aspect of nominal group lexical cohesion, in Coulthard, M. (ed.), *Advances in Written Text Analysis*, London: Routledge.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*, London: Edward Arnold.

Halliday, M. A. K. (1989). *Text, Context, and Learning*, Oxford: Oxford University Press.

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

Ikegami, Y. (1980). "Linguistic typology and textual cohesion: Some notes on Halliday and Hasan's cohesion in English", *Sophia Linguistica*, 6, pp. 92-105.

Johns, A. M. (1980). "Cohesion in written business discourse: Some contrasts", *The ESP Journal,* 1(1), pp. 35-44.

Kai, J. (2008). Lexical cohesion patterns in NS and NNS dissertation abstracts in applied linguistics: A comparative study. Retrieved Dec. 7, 2009, from http://www.linguistics- journal.com/December_2008-ik.php.

Linnarud, M. (1987). "The Swedish project: Lexical chains and lexical cohesion", *Nordwrite Reports II. Trondheim Papers in Applied Linguistics*, pp. 29-39.

MacMillan, F. (2007). "The role of lexical cohesion in the assessment of EFL reading proficiency", *Arizona Working Papers in SLA and Teaching*, 14, pp. 75-93.

Mirshekaran, I. (2003). *Lexical Cohesion in Persian Narrative and Expository Text Types*, Unpublished MA thesis, Shriaz University, Iran.

Mohebbi, B. (1998). *A Contrastive Analysis of Ellipsis in English Texts and their Translation Equivalents in Persian*, Unpublished MA thesis, Shiraz University, Iran.

Morris, J. and Hirst, G. (1991). "Lexical cohesion computed by thesaural relations as an indicator of the structure of text", *Computational Lingustics*, 17 (1), pp. 21-48.

Morris, J. and Hirst, G. (2003). "Term relationships and their contribution to text semantics and information literacy through lexical cohesion", Proceedings of the 31st Annual Conference of the Canadian Association for Information Science. Halifax, Nova Scotia.

Nourmohammadi, E. (1988). Contrastive Analylsis of Cohesion in English and Persian, Unpublished MA thesis, Shiraz University, Iran

Saharagard, R. (1992). "Lexical Cohesion in English and Persian", paper presented at Tabriz- Baku 1st Conference on Language Teaching Issues, Tabriz. Iran.

Seddigh, F. and Yarmohammadi, L. (1996). "Lexical Cohesion in English and Persian in Contrast", *Indian Journal of Applied Linguistics*, 22 (1).

Taboada, M. T. (2000). Collaboration Through Talk: the Interactive Construction of Task Oriented Dialogue in English and Spanish, Unpublished doctoral thesis. Madrid University, Spain.

Tanskanen, S. (1995). "Continuity and interaction: Aspects of lexical cohesion in conversation", *Anglicana-Turkuensia*, 10, pp. 99-114.

Veisi, A. (2002). Lexical Cohesion in English and Kurdish in Contrast, Unpublished MA thesis, Shiraz University, Iran.

Ziahosseiny, S. M. (1999). *A Contrastive Analysis of Persian and English and Error Analysis,* Tehran: Vira Publication.

Appendix