

Study on Lexical Cohesion in English and Persian Research Articles (A Comparative Study)

Fatemeh Mirzapour

Islamic Azad University, Hamedan Branch, Iran

Maryam Ahmadi

Islamic Azad University, Hamedan Branch, Iran

Received: June 7, 2011

Accepted: July 6, 2011

Published: December 1, 2011

doi:10.5539/elt.v4n4p245

URL: <http://dx.doi.org/10.5539/elt.v4n4p245>

Abstract

The present study aims to analyze comparatively English and Persian research articles (Linguistics, Literature, and Library and Information disciplines) in terms of number and degree of utilization of sub-types of lexical cohesion in order to appreciate textualization processes in the two languages concerned. The study analyzes 60 research articles (30 articles in each language) in terms of sub-types of lexical cohesion. The study reveals that the order of occurrence in descending order of sub-types of lexical cohesion is (Rep., Col., Syn., Gen.N., Mer., Hyp., and Ant.) in English data, while the order in Persian data is (Rep., Syn., Col., Ant., Hyp., Mer., Gen.N.). In both data the most frequent sub-types are repetition, collocation, synonymy. In English data the general tendency is towards the use of repetition and collocation but Persian data show the general tendency towards the use of repetition and synonymy. This study might have implications for teachers and researchers in the field of teaching English as a foreign language because of the fact that teaching sub-types of lexical cohesion to foreign language learners will improve the quality of their reading and writing.

Keywords: Lexical Cohesion, English, Persian, Research articles, Comparative study

1. Introduction

In the early seventies, when text analysis was still in its early stages, a number of important works were published dealing with the term cohesion. The most widely known study was that of Halliday and Hasan (1976) in which the devices available in English for linking sentences to each other were classified into reference, ellipsis, substitution, conjunction, and lexical cohesion.

According to Halliday and Hasan (1976), Halliday (1985) and Hasan (1984), the type, number, and degree of utilization of cohesive devices used in the text contribute to the cohesiveness of a text. In spoken and written English discourses, accordingly, individual clauses and utterances are linked semantically by grammatical connections (McCarthy, 1991), which make a text cohesive. For Hoey (1991, p.260), "cohesion is a property of a text whereby certain grammatical or lexical features of the sentences of the text connect them to other sentences in the text."

Cohesion is a semantic concept and it refers to relations of meanings that exist within the text and that define it as a text. So cohesion helps to create text by providing texture. According to Halliday and Hasan (1976), the primary factor of whether a set of sentences do or do not constitute a text depends on cohesive relationships between and within the sentences which create texture: "A text has texture and this is what distinguishes it from something that is not a text... The texture is provided by the cohesive RELATION (1976, p.2). Cohesive relationships within a text are set up "where the INTERPRETATION of some elements in the discourse is dependent on that of another. The one PRESUPPOSES the other in the sense that it cannot be effectively decoded except by recourse to it" (1976, p.4). Consequently a relation of cohesion is set up and presupposed and presupposing elements are integrated into a text. The presupposition and the fact that it is resolved provide cohesion between sentences and the create text.

Malmkjær (2004, p.543) is of the opinion that "cohesion concerns the way in which the linguistic items of which a text is composed are meaningfully connected to each other in a sequence on the basis of the grammatical rules of the language, and formal devices signal the relationship between sentences. Cohesion is a necessary though not a sufficient condition for the creation of the text. The textual or text-forming component of linguistic system of which cohesion is one part, creates text. According to Lotfipour-Saedi (1991) is also of the opinion that cohesion is one of the textual features which makes the texture of a text and helps to its materialization. Cohesion connects certain

grammatical or lexical features of the sentences to the text of the other sentences in the text. Campbell (1994) argues that there are two major principles of cohesive elements by which the continuity aspect of coherence can be explained: 1. the cohesive principle of similarity, 2. the cohesive principle of proximity. The discourse producers influence recipient's sense of discourse continuity by manipulating the similarity and proximity of the full range of discourse elements. The cohesive principle of similarity acknowledges the cohesive effect of similar discourse elements, while the cohesive principle of proximity acknowledges the effect of the spatial and temporal proximity of discourse element. This latter principle acknowledges cohesive effect of deictic discourse elements.

Bex (1996, p.91) considers cohesion "as residing in the semantic and grammatical properties of language. Cohesion guides the ways in which units of text are to be understood in relation to each other. Cohesion concerns the ways in which texts can refer to themselves and is typically achieved through the use of grammatical devices and lexical repetition."

Halliday and Hasan (1976, p.14) argue that cohesion is expressed partly through the grammar and partly through the vocabulary, hence grammatical cohesion and lexical cohesion. "It is necessary to consider that cohesion is a semantic relation but, like all the components of semantic system, it is realized through the lexicogrammatical system. The lexicogrammatical system includes both grammar and vocabulary. Of the cohesive types reference, substitution, and ellipsis are grammatical; lexical cohesion is lexical; and finally conjunction is on the borderline of the two, mainly grammatical, but with a lexical component in it" (Halliday and Hasan 1976. p.5).

The text is not a structural unit and cohesion is not a structural relation. The relation of the parts of a text is not the same as structural relation of the parts of sentences. In other words, cohesion is the non-structural resources for establishing relations within the text to construct discourse. These relations may involve elements of any extent, from single words to a lengthy passage of text. Cohesive ties between sentences are the only source of texture while within the sentences there are structural relations. It is these intersentential cohesion that is important for the text. Within sentence relations since they hang together already, cohesion is not needed to make them hang together.

Cohesion expresses the continuity that exist between one part of the text and another one. This continuity is significant from two aspects. On the one hand, that continuity shows at each stage in the discourse the points of relations or contact with what has been said before. On the other hand, the continuity provided by the cohesion helps the readers to fill in the gap in the discourse, to supply all the components of the message which are not present in the text but are important and necessary to its interpretation. There are some holes in a complete text because it is not possible for the writer to supply all the details. But the reader can supply the missing points even though the text is not complete. It is so because the cohesion makes the interaction between reader and the text possible. Cohesion is used by both readers and writers to create coherence in the text. On the whole, cohesive devices contribute to texture, readability and comprehensibility of a text.

There are five major types of cohesive devices: 1) reference, 2) substitution, 3) ellipsis, 4) conjunction and 5) lexical cohesion. The first four are grammatical and the last one is lexical. According to Halliday and Hasan (1976), lexical cohesion is 'phoric' relation which is established through the structure of vocabulary, and it is a relation on the lexicogrammatical level. Lexical cohesion comes about through the using of items that are related in some way to those that have gone before. In short, lexical cohesion occurs when two words in a text are related in terms of their meaning. Reiteration and collocation are the two major types of lexical cohesion. Reiteration includes repetition, synonymy or near-synonymy, hyponymy (specific-general), meronymy (part- whole), antonymy and general nouns.

1.1 Repetition

Repetition of a lexical item is the most form of lexical cohesion; e.g. **dog** in Reza saw a **dog**. The **dog** was wounded by the children.

In order for a lexical item to be recognized as repeated it need not be in the same morphological shape.

Ali **arrived** yesterday. His **arrival** made his mother happy.

Arrived, arriving, and arrival are all the same item, and occurrence of any one constitutes a repetition of any of the others. Inflectional and derivational variants are also as the same item.

1.2 Synonymy

Lexical cohesion is also created by the selection of a lexical item that is in some sense synonymous with a preceding one.

What **people** want from the government is frankness.

They should explain everything to the **public**.

1.3 Hyponymy (Specific – General)

Hyponymy is a relationship between two words, in which the meaning of one of the words includes the meaning of the other word. For example, the words, animal and dog are related in such a way that dog refers to a type of animal, and animal is a general term that includes dog as well as other types of animal.

A **dog** is a symbol of loyalty. That **animal** is mine.

1.4 Meronymy (Part – Whole)

In this kind of lexical cohesion, cohesion results from the choice of a lexical item that is in some sense in part-whole relationship with a preceding lexical item.

An English **daily** Monday talked about the result of presidential election.

The **editorial** described that pre-election speeches caused good results.

1.5 Antonymy

In this type of lexical cohesion, cohesion comes about by the selection of an item which is opposite in meaning to a preceding lexical item.

Ali **received** a letter from bank yesterday. He will **send** answer next day.

1.6 General Nouns

The general nouns including thing, person, do,... are used cohesively when they have the same referent as whatever they are presupposing.

Saddam doesn't approve **military action against Iraq**. He said that **the moves** was illegal.

1.7 Collocation

This type of lexical cohesion results from the association of lexical items that regularly co-occur. Or as Yarmohammadi (1995, p.127) believes collocation is achieved "through the association of lexical items that regularly tend to appear in similar environments. Such words don't have any semantic relationship". Behnam (1996, p.142) considers collocation as "collocation is one of the factors on which we build our expectations of what is to come next." An example of collocation is as the following:

A huge oil boat polluted the **sea**. Many dead fishes lie along the **beach**.

Hoey (1991) argues that lexical cohesion is the single most important form of cohesion, accounting for something like forty percent of cohesive ties in texts. He continues that various lexical relationships between the different sentences making up a text provide a measure of the cohesiveness of the text. The centrality and importance to the text of any particular sentence within the text will be determined by the number of lexical connections that sentence has to other sentences in the text.

Research Question:

Although different languages use different cohesive devices in creating texts, some specific types of them are used with different degree in different texts. These cases of use can be clarified by cohesion analysis. This study is devoted to lexical cohesion about which there is not enough research. Lexical cohesion, as Yarmohammadi (1995) put it, "is a relation that exist between or among specific sentences in a text and is achieved through the vocabulary." Thus continuity of lexical meaning leads to the establishment of lexical cohesion. So the problem of this research is stated as follows:

- 1) What are the differences between English and Persian research articles in terms of type, number, and degree of utilization of lexical cohesive device?
- 2) What are the similarities between English and Persian research articles in terms of type, number, and degree of utilization of lexical cohesive device?

2. Method

2.1 Sources of data

The data for this study consist of sixty English and Persian research articles, i.e., thirty articles in each language. Quarterly journals of Foreign Language Research, Language and Linguistics, Library and Information, Translation Studies, Research and Planning in Higher Education, and two quarterly journals of Language and Persian Literature were selected Persian journals. Online Information Review, Language Learning, Modern Language Learning, Education and Training, Linguistics and Education, and Renaissance Studies were selected English journals. In order to make the analysis manageable, the articles in the field of language and linguistics, language and literature,

library and information in each language were selected.

To have an almost equal amount of data in English and Persian, the first 250 words from each text were analyzed. The total number of words analyzed amounted to about 15000 in number, 7500 in each language.

2.2 Procedures of data analysis

To analyse the data, first every sentence in each text, and the number of cohesive ties will be detected. Second, the presupposing elements in the cohesive ties will be found. Third, each tie will be specified for the type of cohesion and its related sub-type. Fourth, distance between the presupposed and the presupposing elements will be determined for each tie. Distance is marked as (0) meaning that the presupposed element occurs in the immediate sentence or the same sentence. In other words, “the presupposition is fulfilled in the immediately preceding sentence” (Behnam 1996, p.134). It is marked as (N) when the presupposed and the presupposing elements occur distant from each other. It is marked as (M) meaning that while the presupposing and the presupposed elements are distant but have one or more intervening sentences that enter into a chain of presupposition. “A presupposed element is interpreted with reference to some sentence earlier but with some intervening instances of the same presupposed item” (Behnam 1996, p.134). Finally, it is marked as both (M) and (N) when a tie is both mediated and remote.

Digits in front of (N) and (M) refer to the number of intervening sentences. Then the presupposed item (Pre.It.) which is related to the presupposing item in a tie in some way is specified.

When the analysis of two sets of data is completed, two tables are needed in order to find the frequency of each type of cohesive ties in each data. One table shows sub-types of lexical cohesion in English data and the other one shows the number of sub-types in Persian data.

Again two summarizing tables are needed to present the percentage of sub-types of lexical cohesion in both data. The following formula is used to calculate the percentage of each sub-type:

The percentage of each sub-type = $\frac{\text{The number of that sub-type}}{\text{total number of words}} \times 100$

This formula is used by Hasan (1984) and Yarmohammadi (1995) in their analysis of cohesive devices.

The following lists of abbreviations are used to demonstrate sub-types of lexical cohesion:

- Repetition (Rep.)
- Synonymy (Syn.)
- Hyponymy (Hyp.)
- Meronymy (Mer.)
- Antonymy (Ant.)
- General Nouns (Gen.N.)
- Collocation (Col.)

Abbreviations which are used in data analysis part are:

- S.No. = Sentence Number
- Co.It. = Cohesive Items
- Dis. = Distance
- Tp. = Type
- Pre.It. = Presupposed Item

After describing the methodology, the next step is to demonstrate the number of lexical cohesion in each data. Two tables in this section are used to show the frequency of sub-types of lexical cohesion in each data. One table shows the number of ties in English data and the other shows the number of ties in Persian data. Each table has nine columns for sub-types: number of texts, number of words, frequency of repetition, frequency of synonymy, frequency of hyponymy, frequency of meronymy, frequency of antonymy, frequency of general nouns, and frequency of collocation. Two summarizing tables are used to present the percentage of lexical cohesion in both languages.

3. Results

3.1 Lexical cohesion in English data

The results of analysis of English data in terms of number of sub-types of lexical cohesion are given through the

table 1 at the end of the paper. The total number of lexical cohesion in English data is (1437) of which (830) are repetition, (102) are synonymy, (34) are hyponymy, (41) are meronymy, (43) are general noun, and (303) are collocation.

3.2 Lexical cohesion in Persian data

The results of analysis of Persian data in terms of number of sub-types of lexical cohesion are given through the table 2 at the end of the paper. The total number of lexical cohesion in Persian data is (1420) of which (990) are repetition, (148) are synonymy, (44) are hyponymy, (39) are meronymy, (48) are antonymy, (29) are general noun, and (122) are collocation.

3.3 Lexical cohesion in English and Persian data

In order to calculate the percentage of sub-types of lexical cohesion in two sample data of English and Persian, the number of each sub-type of lexical cohesion represented in tables 1 and 2 and the above mentioned percentage formula are used. So the findings can be represented as: in English data repetition is the most frequently used sub-type of lexical cohesion. The next frequent sub-type is collocation, followed by synonymy, general noun, meronymy, hyponymy, and antonymy. The order is Rep., Col., Syn., Gen.N., Mer., Hyp.

In Persian data also the highest percentage of occurrence belong to repetition. And synonymy, collocation, antonymy, hyponymy, meronymy, general noun show lower percentage of occurrence respectively. The order is Rep., Syn., Col., Ant., Hyp., Mer., Gen.N.

Both data exhibit a general tendency toward the use of repetition, but the average is higher for Persian data (11.06 vs. 13.2).

The occurrence of synonymy, hyponymy, meronymy, antonymy, and general noun is almost the same in both data (1.36 , 0.45 , 0.54 , 0.45 , 0.57 vs. 1.97 , 0.58 , 0.52 , 0.64 , 0.38). Collocation is used more frequently in English data than Persian ones (4.04 vs. 1.62).

4. Conclusion

Two sets of data were analysed in terms of sub-types of lexical cohesion and the results were represented in tables. From the comparative study of lexical cohesion in two sets of data the following conclusions were drawn.

- 1) In terms of sub-types of lexical cohesion the order of occurrence in descending order is (Rep., Col., Syn., Gen.N., Mer., Hyp., and Ant.) in English data , while the order in Persian data is (Rep., Syn., Col., Ant., Hyp., Mer., and Gen.N.).
- 2) In both data the most frequent sub-types are repetition, collocation, and synonymy.
- 3) In English data the general tendency is towards the use of repetition and collocation, but Persian data shows the general tendency towards the use of repetition and synonymy.

5. Pedagogical Implications

Reading is a process of interaction between the reader and the text in which the reader gets meaning from the text but not from isolated sentences. The fact is that there is a difference between a collection of unrelated sentences and a series of sentences comprising a text. This difference can be explained by the existence of some relationships between sentences including theme / rheme, information structure, cohesive patterns...

As elaborated by Yarmohammadi (1995), if a pattern of cohesion becomes evident while analyzing these relationships, it must be that this pattern is at least one factor in the explanation of the greater meaning of a whole text.

So for the EFL and ESP learners knowing the fact that the sub-types of cohesive relations exist within different texts in different order and with different degree of utilization makes the interaction between them and the text easy. Neglecting this pattern (cohesion) is one of the reasons that many Iranian students can not read and comprehend the text outside the class because reading is not treated as it is by Iranian teachers.

The same is true for the students' writing skill. Many students who have graduated from high school can not write a coherent paragraph, even though they can write correct sentences in isolation. But a coherent text not isolated sentences is frequently used. The issue that students can not communicate via written language can be explained by the assumption that sentence elements which create cohesion have not been taught. We should bear in mind that good writers are usually good readers.

References

Behnam, B. (1996). A Stylistic Study of Cohesive Features in English Prose Fiction with Some Pedagogical

- Implications for Non-Native Contexts. A Ph.D thesis submitted to the University of Sheffield, England.
- Bex, T. (1996). *Variety in Written English: Texts in Society, Societies in Text*. London: Routledge.
- Bhatia, K. (1993). *Analyzing Genre-Language Use in Professional Settings: Approach to Genre Analysis*. London: Longman.
- Campbell, K. S. (1994). *Coherence, Continuity and Cohesion: Theoretical Foundations for Document Design*. Denton, TX: University of North Texas Press.
- Cook, G. (1989). *Discourse*. Oxford: Oxford University Press.
- Halliday, M. A. K. (1985). *An Introduction of Functional Grammar*. London: Edward Arnold.
- Halliday, M. A. K., & R. Hasan. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. A. K., & R.Hasan.(1989). *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Oxford: Oxford University Press.
- Hasan, R. (1984). *Coherence and Cohesion Harmony*. Delaware: International Reading Association.
- Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Lotfipour-Saedi, K. (1991). Analyzing Literary Discourse: Implication for Literary Translation. Paper presented at Tabriz University Conference on Translation, Tabriz.
- Malkjar, K. (2004). *The Linguistic Encyclopedia*. London: Routledge.
- McCarthy, M. (1991). *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- Sahragard, R. (1991). *Lexical Cohesion in English and Persian*. Paper presented at Tabriz-Baku 1st Conference on Language Teaching Issues, Tabriz.
- Seddigh, F. (1991). Lexical Cohesion in English and Persian Journalistic Texts: A Contrastive Approach. *Unpublished MA thesis*. Shiraz University.
- Widdowson, H. G. (1973). An Applied Linguistics Approach to Discourse Analysis. *PhD Thesis*. University of Edinburgh.
- Yarmohammadi, L. (1995). Fifteen Articles in Contrastive Linguistics and the Structure of Persian: Grammar. *Text and Discourse*. Rahnama Publishers.
- Yule, G. (1985). *The Study of Language*. London: Cambridge University Press.

Acknowledgement

This article is extracted from the study supported by Islamic Azad University, Hamedan Branch.

Table 1. Number of Sub-types of Lexical cohesion in English Data

No. of Text	No. of Words	Rep.	Syn.	Hyp.	Mer.	Ant.	Gen.N.	Col.
1	250	53	5	-	2	4	1	4
2	250	31	2	-	3	-	2	5
3	250	50	6	4	2	-	-	5
4	250	28	3	4	1	2	1	19
5	250	35	4	-	3	-	-	10
6	250	30	7	-	2	-	-	-
7	250	33	5	-	1	2	1	5
8	250	43	6	7	1	-	-	15
9	250	33	3	2	-	1	1	17

10	250	38	3	-	-	-	-	5
11	250	22	11	-	-	2	4	8
12	250	7	3	-	2	1	1	6
13	250	20	2	-	-	-	2	7
14	250	12	2	3	-	1	1	1
15	250	9	-	2	2	3	1	9
16	250	28	6	-	-	4	2	4
17	250	18	2	1	2	1	-	1
18	250	15	1	-	2	1	-	3
19	250	13	1	9	-	-	3	6
20	250	11	-	-	-	1	1	1
21	250	26	2	1	1	1	-	5
22	250	33	2	-	-	2	7	7
23	250	29	4	-	8	2	2	2
24	250	37	1	1	2	-	2	1
25	250	30	2	-	1	1	1	19
26	250	26	2	-	4	2	5	-
27	250	28	7	-	1	-	1	20
28	250	20	4	8	-	-	4	7
29	250	31	2	-	1	3	-	2
30	250	41	4	1	-	-	-	2
total	7500	830	102	34	41	34	43	303

Table 2. Number of Sub-types of Lexical Cohesion in Persian Data

No. of Texts	No. of Words	Rep.	Syn.	Hyp.	Mer.	Ant.	Gen.N.	Col.
1	250	38	14	1	4	1	-	12
2	250	28	5	-	3	-	2	3
3	250	14	7	-	1	-	1	1
4	250	34	6	-	-	2	1	4
5	250	40	9	2	1	2	-	7
6	250	36	10	4	1	1	-	4
7	250	34	2	1	-	-	7	3
8	250	20	5	-	-	-	4	-
9	250	31	3	-	-	-	-	5
10	250	45	-	-	-	7	-	1
11	250	22	2	1	-	1	3	3
12	250	26	1	-	-	-	-	4
13	250	36	4	1	-	-	-	5
14	250	25	12	-	-	2	1	2
15	250	20	2	1	-	1	1	7
16	250	31	7	-	-	2	-	5
17	250	32	2	2	-	3	2	2
18	250	32	2	3	-	1	-	5
19	250	42	4	-	-	3	2	8
20	250	37	5	7	-	1	-	3
21	250	33	9	4	-	-	1	5
22	250	39	3	-	-	4	-	12
23	250	32	5	7	-	4	-	2
24	250	32	6	4	1	-	1	2
25	250	44	3	-	9	7	-	3
26	250	47	2	1	3	1	-	5
27	250	29	4	3	7	-	1	6
28	250	44	3	2	2	-	1	9
29	250	28	7	-	1	5	-	9
30	250	39	3	-	6	-	4	1
total	7500	990	148	44	39	48	29	122

Table 3. Frequency of Lexical Sub-types in Two Sets of Data

Types of Lexical Cohesion	Rep.	Syn.	Hyp.	Mer.	Ant.	Gen.N.	Col.
English Data	11.06	1.36	0.45	0.54	0.45	0.57	4.04
Persian Data	13.2	1.97	0.58	0.52	0.64	0.38	1.62