

Lexical Cohesion Patterns in Research Articles: Hard Science vs. Soft Science Disciplines

By

¹Mohsen Shahrokhi, ²Aghdas Sadeghi, ²MasoudAmiri Dehnoo

¹English Department, Shahreza Branch, Islamic Azad University, Isfahan, Iran

²Aleshtar Branch, Islamic Azad University, Lorestan, Iran

Abstract

There are many features that contribute to the failure or success in the publication of academic research articles. Thematization, rhetorical moves, information structure, lexis and content are among such features. Lexical cohesion pattern which is one of the most important cohesive mechanisms for the actualization of textual coherences is one of the features in this regard that seems to be unexplored. Drawing upon Hoey's model (1991), this study aims at exploring the way lexical cohesion patterns are distributed in Introduction section of hard and soft science research articles. In so doing, a corpus containing 90 research articles from three disciplines from hard science articles (namely, Chemistry, Physics, and Engineering) and three disciplines from soft science articles (namely, Psychology, Sociology, and Linguistics) were analyzed. For the purpose of our analysis, the frequency and distribution of lexical cohesion patterns were summed; and the Chi-square statistical procedure revealed that there is a significant difference in distribution of lexical cohesion patterns in hard and soft disciplines.

Keywords: Lexical cohesion patterns; hard disciplines; soft disciplines; Chi square

1. Introduction

Within the realm of scientific publication, research articles play an important role in academic communities. Successful publication certainly can lead to reputation enhancement, prestige, research grants, and career advancement. Accordingly, writing for publication has become an essential skill for scientists and researchers across disciplines. There are many features contributing to the failure or success in the publication of academic research articles. Thematization, rhetorical moves, information structure, lexis and content are among such features. Lexical cohesion pattern, among others, is one of the features that still calls for further investigation.

The significance of lexical cohesion and cohesive devices in written and spoken discourse has been emphasized by many researchers (e.g. Johnston, 1987; Sardinha, 1997; Teich & Fankhauser, 2005; Majica, 2006; Klebanov & Shamir, 2007). Halliday and Hassan (1976) presented taxonomy of cohesive devices or ties for English in which they list five main categories: reference, substitution, ellipsis, conjunction and lexical reiteration. The last category, which equates with repetition, refers to a broad range of relation between an item with other preceding words, where the preceding word can be an exact repetition of the first word, a general word, its synonymy or its superordinate. Halliday and Hasan (1976) take lexical cohesion as the central device for making texts interpretable i.e., defining the aboutness of the text. However, Hoey (1991) argues that they fail to note that lexical cohesion is the most important form of cohesive ties. He notes that lexical cohesion is the dominant mode of creating texture. According to Hoey (1991), one of the applications of lexical cohesion patterns is their contribution to the coherence of the text and around forty to fifty percent of cohesive ties of a text are lexical. He proposed a model which has been used by many researchers for exploring the cohesiveness of texts of any genre.

Hoey (1991) in his work, *patterns of lexis in texts*, described different forms of repetition. He introduced in his work a number of potentially computable notions such as links, ties, bonds, and bond networks in relation to lexical cohesion and to text organization. He pointed out that lexical items are semantically

related to other lexical items in the text; hence, they form a network. In other words, he believes that text cohesion is formed by links between words as well as semantic relations between sentences. When two sentences contain an average number of links (usually three or more) a bond is formed between those sentences. For Hoey (1991), the number of links between sentences indicates the significance of the semantic parallelism, which may be either overt or underlying between the sentences. In his proposed model for patterns of lexical cohesion, Hoey (1991) introduces ten categories, which are numbered in decreasing order of importance.

1. Simple lexical repetition
2. Complex lexical repetition
3. Simple mutual paraphrase or simple paraphrase
4. Simple partial paraphrase or complex paraphrase
5. Antonymous complex paraphrase
6. Other complex paraphrase (or Superordinate and Hyponymy)

In addition to those lexical links, however there are other types of textual connections that serve the same function.

7. Substitution
8. Co-reference
9. Ellipsis
10. Dexis

Hoey(1991) believes the most important factor in coherence of a text is the repetition of words along the whole text. Repetition of words also helps the researchers not to lose their way and do their best to just develop the aimed topic (Khani&Tazic, 2010). The role of lexical cohesion patterns in the coherence of texts has been investigated by some scholars. Most of the studies that have already examined the role lexical cohesion patterns have focused on the different types of texts such as narrative (Fox, 1987) expository (Britton, 1994), persuasive (Berzlanovich et al., 2008), and dialogic texts (Buitkiene, 2005). In a study on university entrance examination in Brazil, Batista (2002) argued that the principles in Hoey's (1991) reading model may be applied to the analysis of multiple-choice reading comprehension tests.

Since the 1990s, the emphasis on research into lexical cohesion has gradually shifted from theoretical exploration to genre-based practical analysis (Kai, 2008). In the literature, the majority of research is based on Halliday&Hassan's framework, and little research has been done based on Hoey's model. Kai (2008), following Hoey's (1991) model, took the genre of dissertation abstracts in the discipline of applied linguistics in his study. In his sample, he selected randomly fifteen abstracts written by native and nonnative speakers of English. He found that native speaker abstracts tended to use more complex repetitions than non-native speaker ones, which had a high rate of using simple repetitions. His study also indicated that the patterning of lexical repetition in the sample texts could take a central place in the organization of text. In another study, MacMillan (2007), by using Hoey's (1991) lexical cohesion patterns model, conducted a study for the aim of exploring the implications of the text-forming function of lexical cohesive patterns in English for the assessment of effective EFL reading comprehension. His findings have suggested that lexical cohesion plays a fundamental role in the construct of reading reflected on the TOEFL test.

The present study, drawing upon Hoey's (1991) model, aims at answering the following questions:

1. Which types of lexical cohesive devices are most frequently distributed in hard and soft science disciplines?
2. Are there any similarities and differences in the patterns of lexical cohesion between hard and soft science disciplines?

Hoey's Lexical Cohesion Patterns

The model suggested by Hoey (1991) consists of ten categories which are ranked in decreasing order of importance. These categories include: Simple repetition, Complex repetition, Simple paraphrase, Complex paraphrase, Antonym, Superordinate and Hyponymy, Substitution, Co-reference, Ellipsis, and Deixis. Simple repetitions involve items which Hoey defines as “formally identical” and share the same morphemes with minimum alternations such as plural nouns, verbs making third person singular, simple past and past participle, as well as gerund verbs. Complex repetition occurs either when two lexical items share lexical morphemes, but not formally identical, or when they are formally identical, but have different grammatical functions. *Assessment* and *working memory* are examples of simple repetition in and *early* and *earliest* are examples of complex repetition in example 1.

Example 1: In the assessment of the working memory guidance of attention, early Rts have been emphasized as the best assessment of automatic effects. For example, Soto et al. based their claims of involuntary effects of working memory items on visual attention on their assessment of the earliest 10% of Rts.

In example 2, another example of complex repetition is provided.

Example 2: Clearly, task complexity does interact with planning; planning appears to have a greater effect on fluency in the case of less complex tasks.

The second category of repetition involves simple and complex paraphrase. Simple paraphrase occurs when one lexical item can be substitute for another without changing meaning. Complex paraphrase occurs when it is not possible to make a straight substitution, but it is possible to define one of the items such that it includes the other. Examples of simple and complex paraphrase are respectively as follows:

Example 3: We then assessed whether predation rates differed between the factors. Finally, we tested whether the variation.....

Example 4: Operational definitions have varied considerably, as will become evident in the subsequent sections of this article. These differences in the operational definitions are problematic as they make comparisons across studies difficult in some instances.

The third category of repetition, Antonymous complex paraphrase, involves repetition of antonymous items either belong to the same or different word class.

Example 5: In the unguided condition, learners were simply given the task, told they would have to perform it, and asked to prepare to do so. In guided planning, their attention was directed to specific aspects of planning.

Superordinate and Hyponymy, as the other categories of repetition, relate to cases when two repeated items are the general words related to the specific ones and vice versa. The examples 6 and 7 indicate these two categories, respectively:

Example 6: Rabbits are selective grazers but when preferred foods are scarce, they might be forced to.....but a significant potentials limitation of diversionary feeding is that some animals....(rabbits and animals are superordinate)

Example 7: Most of the research on the damage caused to crops by foraging rabbits has occurred in..... a vast proportion of agricultural lands is devoted to vineyard and olive groves.

Co-reference, as a context- dependent link, occurs when items have the same referent (example 8).

Example 8: This is not to say the interests of children are the same, or that their views should be afforded any special status, but that, like any other citizen, they are entitled to be heard and taken seriously.

The other category includes relations that are neither strict lexical repetition nor paraphrase but that involve cohesive bonds between two items. The specific types are daxis, ellipsis, and substitution. Daxis includes demonstrative pronouns but generally not demonstrative adjectives, because the noun being modified has usually already formed a link (example 9).

Example 9: Tajima (2003) reported that those students in his study who held a positive attitude towards the planning performed more fluently than the non-planners, whereas those that held negative attitudes did not.

Hoey accepts ellipsis (example 11) only in instances where inclusion of the elided item is necessary make the sentence grammatical, whereas substitution includes items like *one*, *do*, clausal *so*, and *other*(example 10).

Example 10: Kawauchi’s (2005) study is an important one because it suggests that any benefit that might be derived from strategic planning is only evident in learners of low or intermediate proficiency, not in advanced learners.

Example 11: There are three other notable methodological problems. The first is that none of the studies reviewed in this article collected baseline data of native speakers performing the task.

The corpus

The corpus of this study consisted of 90 Introduction sections of research articles from three hard science disciplines (Physics, Chemistry, Engineering) and three soft science disciplines (Psychology, Sociology, Linguistics). The reason for selecting these two corpora is grounded in Becher and Trowler’s(2001) classification of disciplinary communities in soft and hard sciences. These authors describe hard sciences as those ones belonging to the scientific, technological, and biomechanical fields and soft sciences as those ones belonging to the social science, education, humanities, and art fields.

The articles were chosen from Journals of ScienceDirect available online. Then, the selected articles were scanned scrupulously. Adopting Hoey’s model, lexical cohesion patterns were manually enumerated in articles.

2. Results

In order to determine the frequency of lexical cohesion patterns in the RA introductions across the six disciplines, the observed frequencies of each pattern were counted and summed. Since our first question was which types of lexical cohesions are most frequently distributed across disciplines, the order of categories frequently occurred across soft disciplines was displayed in Table 1 as follows: simple repetition, complex repetition, simple paraphrase, complex paraphrase, co-reference, antonym, superordinate and hyponymy, substitution, daxis, ellipsis were identified most frequently respectively. As for hard disciplines, the order of frequency was as follows: simple repetition, complex repetition, simple paraphrase, complex paraphrase, antonym, co-reference, daxis, superordinate and hyponymy, substitution, and ellipsis.

Table 1: lexical cohesion patterns in hard and soft science articles in order of frequency

	Simple Repetition (%)	Complex Repetition	Simple Paraphrase	Complex Paraphrase	Co-reference	Antonym	Super-ordinate and Hyponymy	Daxis	Substitution	Ellipsis
Soft Science	62.641	15.093	10.423	3.935	2.135	2.051	1.1810	1.050	1.000	0.483
	Simple Repetition (%)	Complex Repetition	Simple Paraphrase	Complex Paraphrase	Antonym	Co-reference	Daxis	Super-ordinate and Hyponymy	Substitution	Ellipsis
Hard Science	70.05	10.526	8.953	2.369	2.222	1.719	1.719	1.132	1.132	0.167

The distribution of lexical cohesion patterns in each of six disciplines is displayed in Table 2.

Table 2: Percentage of lexical cohesion patterns in hard and soft science disciplines

	Simple Repetition (%)	Complex Repetition	Simple Paraphrase	Complex Paraphrase	Co-reference	Antonym	Superordinate and Hyponymy	Dexis	Substitution	Ellipses
Hard Sciences	70.05	10.526	8.953	2.369	2.222	1.719	1.719	1.13	1.132	0.167
Engineering	23.611	2.903	2.822	0.772	0.750	0.571	0.571	0.387	0.388	0.052
Physics	23.481	3.600	2.931	0.776	0.732	0.572	0.571	0.372	0.371	0.057
Chemistry	22.961	3.723	3.200	0.781	0.740	0.576	0.577	0.372	0.371	0.058
Soft Sciences	62.641	15.093	10.423	3.935	2.135	2.051	1.1810	1.05	1.000	0.483
Linguistics	22.85	5.00	2.964	0.902	0.708	0.611	0.370	0.34	0.320	0.102
Psychology	19.09	5.20	3.832	1.300	0.701	0.630	0.313	0.34	0.302	0.150
Sociology	19.94	5.11	3.678	1.110	0.703	0.711	0.311	0.31	0.311	0.110

The second question was related to the similarities and differences observed in the distribution of lexical patterns between the two corpora. As displayed in Table 1, the main similarity between the two corpora lies in the frequency of simple lexical repetition that enjoys a higher frequency compared to other patterns across disciplines. The number in hard disciplines reaches 70 % whereas in soft disciplines it amounts to 62%, which is more numerous than the total of the other patterns of lexis.

Figure 1 provides a visual comparison of the frequency distribution of the lexical cohesion patterns identified across soft and hard disciplines.

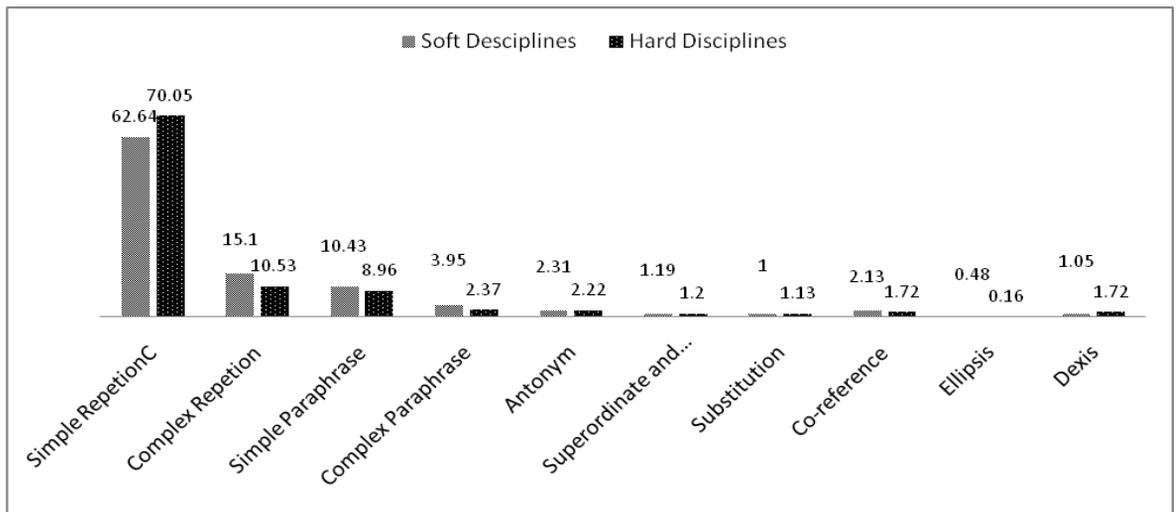


Figure 1: The comparison of frequency of lexical cohesion patterns in soft and hard disciplines

In addition, there are few instances of superordinate and hyponymy, substitution, co-reference, dexis, and especially ellipsis in both groups. As displayed in Figure 1, the distribution of the categories of lexical cohesion patterns is roughly the same in two groups. However, the percentage of categories is different between the groups. For example the first four categories occurred in the two corpora are simple and complex repetition and simple and complex paraphrases, however, the percentage of occurrence of

categories are about 62%, 15%, 10%, and 2.9% respectively for soft science disciplines and 70%,10%, 8%, and 2.36% for hard science disciplines. Furthermore, as displayed in Table1, co-referenceis ranked in fifth place in soft disciplines, whereas in hard disciplines antonym is ranked in this place. In hard disciplines dexis enjoys higher frequency than soft disciplines.

Then, a chi-square test was performed to find out the significance of frequencies obtained in the two corpora.

Table 3: Chi-SquareTests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.135E2 ^a	9	.000
Likelihood Ratio	115.631	9	.000
Linear-by-Linear Association	13.028	1	.000
N of Valid Cases	10765		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 16.39

As Table 3 presents, the distribution difference of lexical cohesion patterns between the introduction section of soft and hard science research articles is statistically significant.

3. Discussion

The main purpose of the present research was to investigate the frequencies of lexical cohesion patterns in three soft science disciplines (namely, Sociology, Psychology, and Linguistics) and three hard science disciplines (namely, Chemistry, Physics, and Engineering).This study also examined the similarities and differences in distribution of lexical patterns in the six disciplines. Results show that simple repetition is the most frequent lexical cohesion pattern identified in both corpora and ellipsis is the least frequently identified pattern.Although similar patterns of use are found in six disciplines concerning lexical patterns, differences are also observed. The percentage of instances of lexical cohesion patterns is different between two corpora. The chi-square test indicates that the difference in the distribution of lexical cohesion patterns between both corporais significant.

Results of descriptive statistics indicated that simple lexical cohesion which is a common phenomenon in English (Zhu, 2001; Miao, 2002; cited in Kiao, 2008) is the most frequent pattern in both corpora. The few instances of substitution, co-references, dexis, and ellipsis reveal that these articles are formal. As displayed in Table 1, differences lie in the percentage of simple repetition, complex repetition, simple paraphrase, and co-reference from hard science to soft science articles, which shows that soft science articles demonstrate a relatively higher level of sophistication than that of hard sciences. Moreover, although the percentages of lexical cohesion patterns are not the same among the disciplines of hard science, the percentage differences do not seem significantly different; this also holds true regarding the disciplines of soft science. In other words, as hard science disciplines are not different regarding the use of different lexical cohesion patterns, the hard science disciplines under in investigation in this study are not different either in terms of the frequency use of lexical cohesion patterns.

However, the difference between the hard and soft sciences can be explained by the different nature of these two disciplines. In fact, since these two sciences are from different genres with particular purposes and audience, writers are required to employ different choices. That is, it is disciplinary which underlies the most specificity (Hyland, 2009).The above-mentioned explanations are further supported by the characteristics Coffin et al. (2003) mentioned for hard and soft sciences. They have described hard

science disciplines as empirical and technological disciplines relying on quantitative and experimental methods, but soft science disciplines as explicitly interpretive disciplines relying on discursive and qualitative methods. Or as argued by Becher and Trowler (2001), the distinction is supported by the different methodological and conceptual frameworks which are used in these disciplines.

Furthermore, since academic texts are about persuasion and this involves making choices to argue in ways which fit the community's assumptions, methods, and knowledge (Hyland, 2009), disciplines such as sociology, psychology, and linguistics give more attention to explicit interpretation than other fields. In other words, academic discourse helps to give identity to a discipline. This means that we need to understand the distinctive ways they have of presenting findings and arguments. Accordingly, it is clear that writers in different disciplines represent themselves, their work and their readers in different ways, with those in the humanities and social sciences taking far more personal positions than those in the sciences and engineering. Successful communication in the soft fields depends far more on the author's ability to invoke the sense of a real writer in the text, emphasizing their own contribution to the field while seeking agreement for it (Hyland, 2004). In the hard sciences, researchers are generally seeking to downplay their personal role in the research to underline the phenomena under study, the replicability of research activities, and the generality of the findings (Hyland, 2009).

In addition, some research evidence indicates that the disciplinary conventions significantly constrain writing style and that sometimes these conventions may have an even greater effect on the writer's choices (Breivegaet al, 2002; Hyland, 2002). According to Hyland (2009), academic writing, much like any other kind of writing, is only effective when writers use conventions that other members of their community find recognizable and convincing. Essentially the process of writing involves creating a text that we assume the reader will recognize and expect.

4. Conclusion and Implication

The aim of the research was to explore lexical cohesion patterns of hard and soft sciences disciplines, and to examine the similarities and differences in the lexical cohesion patterns between hard and soft science articles. According to the results obtained, the frequency use of lexical cohesion patterns between hard science and soft science disciplines was significantly different; however, as hard science disciplines were not observationally different significantly from one another in terms of lexical cohesion patterns, the soft science disciplines did not seem to be different from one another significantly.

According to the results, simple repetition is the most frequent lexical cohesion pattern used by both corpora, and ellipsis is the least frequently used pattern. Although similar patterns of use were found in the six disciplines concerning lexical patterns, there were also differences. The percentage of instances of lexical cohesion patterns is different between the two corpora. The differences can be attributed to the different nature of these disciplines. Hard and soft sciences possess distinctive characteristics which influence the choices the writers make to apply lexis.

Indeed, the fact that writers in different fields draw on different resources to develop their arguments, establish their credibility and persuade their readers means that EAP teachers need to take the disciplines of their students, and the ways these disciplines create texts, into account in their classroom practices (Hyland, 2002). Such considerations, moreover, are not confined to student writing but underlie the conventions which guide scholarly writing for academic publication. The value of discourse analysis is not that it merely produces a list of the features of disciplinary discourses, but that it can uncover more sophisticated understanding of disciplinary communities. It provides a richer picture for academics and for teachers of EAP and so helps language teachers to improve the ways they prepare their students for their academic studies (Hyland, 2002).

Furthermore, textual research can provide the kind of knowledge that can help teachers become familiar with the nature of the reading and writing that their students are required to do in college classrooms, and that can contribute to the effectiveness of articles student write for publication. If composition teachers are to collaborate successfully with teachers in other disciplines on designing ways to use writing as a medium for learning, then knowledge of the similarities and differences of writing in various disciplines and their degrees of importance in distinguishing the writing of a particular discipline should be of high value. In addition, research in discourse analysis has greatly promoted our understanding of specificity, and how language varies in different contexts, and it has become an invaluable tool for scholars and teachers, highlighting typical patterning and salient features of academic writing (Hyland, 2009).

References

- Batista, F. M. (2002). *Investigating reading strategies using think-aloud protocols*. Unpublished M.A. dissertation, Fluminense Federal University.
- Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. Buckingham and Philadelphia: The Society for Research into Higher Education and Open Up.
- Berzlanovich, I., Egg, M., & Redeker, G. (2008). Coherence structure and lexical cohesion in expository and persuasive texts. In A. Benz, P. Kuhnlein & Stese (Eds.), *proceeding of the workshop constraint in discourse* (pp. 19-26). Potsdam: Germany.
- Britton, K. B. (1994). Understanding expository text. Building mental structures to include insights. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 461-674). San Diego: Academic Press.
- Buitkiene, J. (2005). Variability of cohesive devices across registers. *Studies About Language*, 7, 17-20.
- Coffin, C., Curry, M., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2003). *Teaching academic writing: A toolkit for hs in academic higher education*. London: Routledge.
- Fox, B. A. (1987). *Discourse structure and anaphora: Written and conversational English*. Cambridge: Cambridge University Press.
- Halliday, M.A.K & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Hyland, K. (2004). Disciplinary interactions: metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13, 133-151.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27 (1), 4-21.
- Hyland, K. (2009). Writing in the disciplines: Research evidence for specificity. *Taiwan International ESP Journal*, 1(1), 5-22.
- Hyland, K. & Tse, P. (2007). Is there an 'academic vocabulary'? *TESOL Quarterly*, 41 (2), 235-254.
- Johnstone, B. (1987). "Perspectives on Repetition." *Text*, 7(3), 205-214.
- Kai, J. (2008). Lexical Cohesion Patterns in NS and NNS Dissertation Abstracts in Applied Linguistics: A Comparative Study. *The Linguistic Journal*, 3(3), 132-159.
- Khany, R. & Tazik, Kh. (2010). *The relationship between rhetorical moves and lexical cohesion patterns; the case of Introduction and Discussion sections of local and international research articles*. Unpublished Master Thesis. University of Ilam, Iran.

- Klebanov, B. B., & Eli Shamir, E. (2007). Reader-based exploration of lexical cohesion. *Language Resources & Evaluation*, 41, 27–44.
- Lovejoy, K.B. (1991). Cohesion and information strategies in academic writing: analysis of passages in three disciplines. *Linguistics and Education*, 3, 315- 343.
- MacMillan, F. (2007). The role of lexical cohesion in the assessment of EFL reading proficiency. *Arizona Working Papers in SLA & Teaching*, 14, 75-93.
- Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes*, 21(1), 1-17.
- Sardinha, B. A. P. (1997). *Automatic identification of segments in written texts*. Unpublished Ph.D. Thesis, AELSU/English Department, University of Liverpool, UK. Available at www.theses.f2s.com.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Teich, E., & Fankhauser, P. (2005). Exploring Lexical Patterns in Text: Lexical Cohesion Analysis with Word Net. *Interdisciplinary Studies on Information Structure*, 2, 129–145.