Cukurova University Faculty of Education Journal
Vol: 41 No: 2 Page:01-17
http://egitim.cu.edu.tr/efdergi

COH-METRIX: INTRODUCTION AND VALIDATION OF AN ONLINE TOOL FOR TEXT ANALYSIS

Yasemin KIRKGÖZ[a*]; İhsan ÜNALDI[b]

[a]Çukurova University, Faculty of Education, English Language Teaching Department, Adana/Turkey
[b]Gaziantep University, Faculty of Education, English Language Teaching Department, Gaziantep/TURKEY

ABSTRACT

This is a corpus-based study which tries to compare lexical networks of Turkish EFL learners with native speakers of English; it also tries to verify the validity of an online database (coh-metrix), which is used for lexical computations of texts written in English. The lexical computations are performed through 60 indices such as syntax, general word and text formation and referential and semantic aspects. Three different corpora (one learner, two native) were employed to test the tool's ability to differentiate learner texts from native ones. The learner text sets were written by 49 intermediate Turkish students learning English, and the other two text sets were written by two different native groups of speakers of English (100 native speakers in total). A number of statistical testing techniques including Kruskal Wallis and Mann Whitney U-test were used. Some of the indices yielded statistically significant differences between three groups; some of them were able to exclude the learner text sets from the native text sets regardless of average number of words in the texts and the prompts for the essays. The results showed that sentences in the essays written by the Turkish EFL learners lacked lexical cohesion when compared to the sentences produced by the native groups.

Key words: *Lexical networks, Coh-metrix, Corpus.*

INTRODUCTION

Second language learning and acquisition (SLL/SLA) is an expanding field with newly emerging sub-fields. This domain is, in fact, a multi-disciplinary one which gleans insights and methods from a range of disciplines such as linguistics, sociology, sociolinguistics, psychology, psycholinguistics and education (Ellis & Barkhuizen, 2005:3). Naturally, being multidisciplinary comes with rapid developments, and these developments are, most of the time, on a par with new technology.

In theory, research possibilities in SLA are vast; however, much of SLA research has traditionally focused on describing learner language or learners' *interlanguage*; their sequences of development have been the focal point in these studies (Pica, 2005). Coined by Selinker in 1972, the term interlanguage or learner language could be defined as the interim stage between a learner's native language (L1) and the target language (L2) s/he is trying to learn.

Meara (2002), in a recent review of four books about second language lexical acquisition, highlights some important issues. First of all, he claims that unlike syntax and morphology, lexical development in L2 has been sidelined for years since 1950's, and he goes on to say that

* **Co-Author:** ykirkgoz@cu.edu.tr

this is a rediscovery period in terms of L2 lexicon, because studies which were once overlooked now make sense. This situation, in fact, creates a problem because now it is time to fill the gap of a sound L2 lexicon theory.

Lexical errors of language learners are, in fact, global errors (Ellis, 1995; Gass & Selinker, 2008), which means that these errors cause communication breakdowns. The main importance of the current study is that it is an attempt to systematically determine problems concerning Turkish EFL learners' lexical networks; as a matter of fact, this alone is an end itself.

Generally speaking, lexical cohesion in learner language is not a no man's land completely, but more studies are needed about the issue. When we look at the issue with Turkish EFL learners in mind, this time we definitely have a no man's land in front of us. Bearing in mind the lack of studies concerning cohesion in Turkish EFL learners, this study could be regarded as the first attempt to deal with the problem.

THEORETICAL BACKGROUND

The characteristics of learner language have been researched from numerous aspects. With contrastive analysis as a paradigm, this stage has been analyzed for lexical and grammatical errors since 1960's, and to some researchers the assumption was that these errors stemmed from an interference of L1 in L2 acquisition process. Contrastive rhetoric, whereby discourse features of L2 are examined, has also been among the research topics in SLA. The outcomes of these studies have been discussed, analyzed, confirmed or denied by researchers; however, some aspects of learner language have been ignored. Among these aspects, lexical cohesion in learner language is a potentially fruitful one. In language teaching-learning context, trying to deal with *cohesion* in learners' texts is like sailing into uncharted waters; traditionally, it lacks attention (Cook, 1989; Flowerdew, 2009). This lack of attention seems to be noteworthy, as the use of lexical cohesive ties has been reported to be a significant differentiating factor between native and non-native speaker writing (Connor, 1984).

If significant differences exist between L1 and L2 written productions, in order to be able to make intelligent differences about adopting and/or adapting L1 practices, ESL practitioners need to have a clear understanding of these differences (Silva, 1993). In order to develop a clear understanding of the nature of L2 writing, in his seminal meta-analysis, Silva (1993) screened and analyzed 72 empirical reports involving a direct comparison of L1 and L2 written productions. The subjects in his study came from different language backgrounds involving L1s like Arabic, Chinese, Japanese and Spanish. They were predominantly undergraduate students in their late teens or early twenties with fairly advanced English proficiency levels. The reports involving these subjects were compared in terms of fluency, accuracy, quality, structure, morphosyntactic/stylistic and lexicosemantic features. The results suggested that, in general, adult L2 writing is distinct from and less effective than L1 writing. L2 composing appears to be more constrained, more difficult and less effective. L2 writers appeared to be doing less planning and having problems with setting goals, generating and organizing materials. Their transcribing was more laborious, less fluent, and less productive. Reviewing, rereading and reflecting were less, but they revised more. Naturally, they were less fluent and less accurate. In terms of lower level linguistic concerns, L2 writers' texts were stylistically distinct and simpler in structure. Their sentences included more but shorter T-units, fewer but longer clauses, more coordination, less subordination, less noun modification, and less passivization. One important point was about the use of cohesive devices. They used more conjunctive and fewer lexical ties, and exhibited less lexical control, variety, and sophistication.

Another similar and important study was carried out by Ferris (1994). A corpus of 160 texts was analyzed. There were 40 texts each by students from four L1 groups: Arabic, Chinese, Japanese and Spanish. The papers were from a university placement exam in which they were asked to write about culture shock. 62 quantitative, lexical and syntactic features of the text were identified and counted in the corpus. For the purpose of statistical analysis, some of these features were either dropped or combined and 28 of them were left. Some of these features were number of words, impersonal pronouns, modals, negation, coordination, coherence features and repetition. The groups were divided into two; one of which consisted of learners at lower level of proficiency in English and the other consisted of advanced learners of English. A discriminant analysis was performed to see how the mentioned variables would discriminate the two groups. The results revealed that students at higher levels of L2 proficiency used a variety of lexical choices, syntactic constructions, and cohesive devices, and their texts received higher holistic scores. The study also showed that micro-level attention and instruction might be of more significance than many practitioners realized.

Hinkel (2002) carried out a large scale empirical analysis of 68 lexical, syntactic and rhetorical features of L2 text. The corpus included texts written by advanced learners of English from six different languages: Arabic, Chinese, Indonesian, Japanese, Korean and Vietnamese. According to Hinkel, even after years of study in English, the learners still lack some aspects that native speakers have. The results of her study indicate that L2 writers have a severely limited lexical and syntactic repertoire. This led the learners to produce simplistic texts, which are rooted in conversational discourse in English language. The results reveal that there appears to be a big gap between L1 and L2 texts in terms of basic academic writing, which requires alternative methodologies in pedagogical applications in teaching writing. Hinkel (2002:74) listed the features with significantly higher median frequency rates in native speaker (NS) and nonnative speaker (NNS) texts as follows:

- Interpretive nouns
- Vague nouns
- Assertive pronouns
- Public verbs
- Private verbs
- Expecting/tentative verbs
- Modal verbs of necessity
- Be as a main verb
- Predicative Adjectives
- Amplifiers
- Other adverbs (manner, conjunct, and adjective/verb modifiers)
- Adverb clauses of cause
- Phrase-level conjunctions
- Sentence-level conjunctions (transitions)
- Exemplification markers (for example)
- Emphatics

With the expansion of digital text analysis techniques, tools and methods, it is now possible to convert any text written in English into numerical values for analysis, and coh-metrix is one of these digital tools. The online database coh-metrix was first introduced by a team of researchers' study (Graesser et. al., 2004) where the indices were detailed, and one of the

indices in coh-metrix tool, LSA, was tested to explore how it can be used as a method to examine lexical development of L2 speakers. The aim of the study was to see if LSA measures of semantic co-referentiality increases as learners study an L2, and to investigate whether a common measurement of lexical proficiency demonstrates growth, as well. A group of L2 English learners who were enrolled in an intensive language learning program at a state university in the United States were involved in the study. Their lexical growth was tracked by making use of LSA over a long period of time. The participants were at the lowest proficiency level at the beginning. A spoken corpus was formed through interviews over one year. The data collected in the 2nd, 4th, 16th, 32nd, 50th and 52nd weeks were computed. Through statistical analysis, the results revealed that the values computed in the last meeting (52nd week) were statistically significant from those of the first meeting. It was concluded that, in time, subjects' proficiency levels increased in terms of the lexical relations in their utterances.

By making use of the indices in coh-metrix, in a recent and comprehensive study Crossley and McNamara (2009) explored how lexical differences, related cohesion and lexical networks, can be used to distinguish between texts written by first language writers of English and second language writers of English. Two corpora were used; one was from LOCNESS (Louvain Corpus of Native English Essays), and the other one comprised essays written by Spanish learners of English taken from the International Corpus of Learners of English (ICLE). The learners' age and their learning contexts were similar: they were all university students in their twenties. The native corpus comprised of 208 texts (151,046 words, in total) and the learner corpus was comprised of 195 essays (124,176 words, in total). Both corpora included argumentative essays whose topics were again taken from ICLE. A discriminant function analysis was conducted; and in the process coh-metrix indices that measure lexical features related to cohesion and lexical networks were selected. The texts were compared in terms of word hypernymy, word polysemy, argument overlap, motion verbs, CELEX written frequency, age of acquisition, locational nouns, LSA givenness, word meaningfulness, and incidence of casual verbs. The results demonstrated that deeper-level lexical indices related to cohesion and network models in coh-metrix tool can significantly distinguish between L1 and L2 texts. The importance of this study is that, as contrast to the related literature (Connor, 1984; Reynolds, 1995, cited in Crossley and McNamara, 2009), it is the first study to distinguish L1 and L2 texts solely based on lexical features.

What is Coh-metrix

Coh-metrix is an online database, which can assess texts in English at multiple levels. While making calculations about texts, it takes into account five indices: *readability scores, general word and text information, syntax, referential and semantic aspects and situation model dimensions.* Each of these indices is composed of several other indices. From this respect, although some counting is done, coh-metrix is not a word counter in classical terms. It is highly analytical, and singles out every aspect of a text from the others by yielding precise numerical values. As the concern of the study is lexical cohesion in EFL learners' texts, only the referential and semantic aspect scores will be analyzed.

Referential and Semantic Aspects in Coh-metrix

This index focuses on referential cohesion i.e. *Coreference*. Referential cohesion is generally a matter of *overlapping* of constituents within a text. Argument overlaps and stem overlaps between adjacent and distant sentences are taken into account in this index. There are other indices such as anaphor reference and latent semantic analysis.

Anaphor reference index refers to the referential tools i.e. pronouns used in a text. This index measures these referential tools taking adjacent references and references occurring up to five sentences, earlier in a text.

Argument overlap is a proportion ratio score, which calculates the ratio of sentence pairs sharing one or more arguments (*nouns, pronouns etc.*).

Stem overlap refers to proportion of adjacent sentences sharing common word stems. For example, in the following sentence;

The students prepared their presentations meticulously. That's why the preparations took weeks.

The words *prepared* and *preparations* share the same stem and in the database it is called a stem overlap.

Another way through which coh-metrix determines similarity is Latent Semantic Analysis (henceforth LSA). LSA, also known as Latent Semantic Indexing or Correspondence Analysis, is a mathematical and statistical technique for representing word knowledge based on a large corpus of texts. It makes use of Singular Value Decomposition (SVD) technique, which could be regarded as a type of *factor analysis* reducing large corpora of texts to much fewer dimensions (See Crossley et al., 2008 for details).

LSA is generally used to put different body of texts into categories. It is not a simplistic word count or co-occurrence estimation, but a deeper (*latent*) level of mathematical analysis of words. This technique is reported to mimic human word sorting (Landauer et. al., 1998).

THE CONTEXT AND THE PROBLEM

The problem of the current study relates to freshman engineering students learning English as a foreign language at the Higher School of Foreign Languages at University of Gaziantep. In this institution, throughout years, language teaching has been modified, modernized and eventually relatively improved. However, lexical cohesion, especially in learners' written productions, is nowhere near adequate. Although discussions concerning the issue go on continuously, let alone trying to come up with feasible solutions, the problems have not been named, yet.

The common view among the teaching staff at this institution is that engineering students do better in grammar subjects, but when it comes to learning and retaining new vocabulary items and using them appropriately, the teaching/learning process falters. This topic is an ongoing one in teachers' rooms. When the learners are asked to talk about their problems they encounter while learning a second language, the very same topic surfaces. When the written productions of these learners are examined, which is done officially during mid-terms and final exams, teachers' observations concerning lexical cohesion are confirmed.

The primary aim of the current study is to determine lexical and cohesive differences between texts written by Turkish EFL learners and texts written by native speakers of English. These differences are expected to shed light onto the lexical and cohesive flaws in learners' texts.

From the theoretical point of view, this study could be regarded as an attempt to answer certain questions, and raise new ones concerning written productions of EFL learners bringing the interlangual lexicon and cohesion in learners' text into the foreground. These questions are,

however, context-bound i.e. they are limited to a certain teaching/learning context. The rationale behind this paradigm is that every learner, every teacher and every teaching/learning context is unique (Brown, 2007); hence, the problems surfacing in any context need to be handled by taking into account the parameters in that same context.

RESEARCH QUESTION

Taking the related literature into account, the following research question is the main concern of this study:

Regardless of prompt or average number of words used in the texts, to what extent do texts written by Turkish EFL learners deviate from texts written by native speakers of English in terms of referential and semantic aspects?

Assumptions and Limitations

It is assumed that individual differences among the subjects participated in the current study, such as socio-economic and cultural backgrounds will not have significant effects on the statistical outcomes.

Computerized analyses of L2 essays in large scale assessments such as the Test of English are reported to have misidentified L2 textual features with an error ratio of 21 % (Frase et al., 1999). The related assumption is that the online tool, coh-metrix, yields reliable measurements concerning both L1 and L2 corpora.

In this study, a collection of texts written by native speakers of English is used as the reference point for comparison with Turkish EFL learners. The assumption, albeit a strong one, concerning this point, is that the reference corpus, collected from native speakers, is flawless in terms of lexicon and grammar, which would mean that the more native-like a text is, the more coherent it is.

As for the limitations, the current study is limited to Turkish EFL learners whose proficiency levels vary from intermediate to advanced. Furthermore, the number of the learners (49) is too limited for broader generalizations.

METHOD

Participants

Initially, the participants were 850 freshman engineering students at a state university in Turkey. Their ages varied from 19 to 23, and most of the participants were male. In order to meet the requirements for a learner corpus (Granger, 2003), subjects' proficiency levels of all 850 students were determined using a valid and reliable placement test (Allen, 1992). The results were checked to see if their levels were homogeneous, as would be expected. However, the results of the test showed that the subjects' levels varied from A2 (elementary) to C2 (advanced) level, which, in our case, demanded adjustments concerning homogeneity. Therefore, only intermediate and upper level subjects (49) were involved in the study. The assumption was that the subjects who have proficiency levels lower than intermediate level would still be dealing with some basic grammar and lexical issues, which was likely to affect the results negatively. From this respect, a purposive sampling was performed.

Corpora

One of the corpora used in this study was an original one, i.e. compiled by the researcher. The texts were written in a writing exam, photocopied the same day, and handed out back to the students the following day. The students were asked to digitalize the texts, and send them back through email. Since the scope of the study was almost entirely lexical, the students were allowed to make spelling corrections in their texts before sending them. They were also asked to fill in a *student profile* for further descriptive analysis.

The selection of corpora was one of the most demanding parts of the current study. The standard view would be the comparison of two different sets of corpora, and then making interpretations about the results. In their studies, Crossley and McNamara (2009) compared two sets of corpora; one native (L1) and the other learner (L2). They concluded that the online database (coh-metrix) they used in their study is able to distinguish between L1 and L2 texts. In this study, another dimension was added to the equation: a third set of L1 texts. Table 1 is a description of the three corpora used in this study.

Table 1. Comparison of the Three Corpora Used in the Study

Name of the Corpora	Total Number of Words	Average Words per Essay	Essay Type	Prompt
Learner corpora (L)	16,334	333	Argumentative	Exam/Timed
Native corpora 1 (N1)	21,605	400	Argumentative	Exam/Timed
Native corpora 2 (N2)	54,397	1182	Argumentative	Untimed

The average number of words for L2 texts is 333, and for the first L1 corpus it is 400 words per text; however, the second native corpus (N2) has an average of 1182 words per text. All three corpora comprised of argumentative essays, and all essay topics were taken from Granger (1993). The rationale for adding another native corpus is that if the mentioned database or software is capable of distinguishing between L1 and L2, then it should not be able to distinguish between two L1 sets of corpora. In order to determine this issue, Kruskal Wallis and Mann Whitnet U-test, one-way ANOVA and Scheffe (post-hoc) tests were conducted with an expectation that the L2 text sets would differ from both L1s.

RESULTS

Normality and Homogeneity of the Data

Since the study involves multiple groups, before beginning to analyze the data, group scores were tested to see if they were suitable for parametric comparisons. It is common knowledge that in order to be able to make use of parametric tests and make inferences regarding their results, the scores gathered from the subjects must be normally distributed. With very large populations normality is generally not a concern. However, if the population is not that large, the assumption that the subjects' scores are distributed equally has to be tested. The next concern about parametric comparisons is the homogeneity of variance of the scores; it is the requirement that the variances should be the same throughout the data. Generally, the data to be used in any study comes from different populations; if the score variances of these groups are homogeneous, then these groups are suitable for parametric comparisons.

In order to test normality and homogeneity of our subjects' scores, frequency measures and Levene's test of homogeneity of variance were performed. The results concerning the three groups, as a whole, (Learner, Native 1 and Native 2) are exhibited in Table 2.

Table 2. Normality and Homogeneity Results for the Three Groups

Name of the index	sd	Skewness	Standard Error	*z*	*Levene test p value*
Referential and Semantic Aspects					
Anaphor reference (Adjacent)	,138	,370	,196	1,887	.028**
Anaphor reference	,080	,935	,196	4,770*	.008**
Argument overlap (Adjacent)	,145	-,412	,196	2,102*	.497
Argument overlap (All)	,133	,516	,196	2,632*	.013**
Adjacent stem overlap	,170	-,131	,196	,668	.032**
Stem overlap (All)	,16	,372	,196	1,897	.032**
LSA sentence adjacent	,057	-,228	,196	1,163	.985
LSA sentence all	,06	,216	,196	1,102	.487

*Values greater than 1.96 are significant at .05 level.
**Significant at .05 level

In Table 2, the first column represents the name of the index from coh-metrix. The second column indicates the standard deviation values. Skewness values are indications of how much the score distributions are *skewed* compared to a perfectly distributed one. The *z* value is the result obtained from the division of Skewness value by the standard error value. If the obtained score from this division is greater than 1.96, which is taken from the normal distribution table, it means that the scores are *not* normally distributed. The last parameter to be checked, Levene's test of homogeneity is given in the last column. This test checks whether the variance of the scores of a given population are homogeneous or not. If these values for a certain index are significant ($p<.05$), then it can be claimed that the groups are not suitable for parametric comparisons.

A quick glance at Table 2 will make it clear that nearly all of the indices violate either normality or homogeneity assumption, or in some cases both assumptions are violated (e.g. anaphor overlap). However, LSA scores, both adjacent and all-distances, appear to have normal distributions and homogenous variances.

Taking all the above analyses into account, as the group scores are not either normally distributed or lack homogeneity, Kruskal Wallis test, a non-parametric test for multiple groups, has been employed in comparison of the three groups. Since the total participants in this part of the study is relatively large ($n_{total}=149$), Monte-Carlo method has been employed to determine the exact significance values for each of the comparison done by using Kruskal Wallis test. Since we have three groups to compare, a post-hoc test is considered necessary to see which of them will be excluded from the group. At this point, carrying out a post-hoc test is not as easy as it is in parametric tests although it is not a dead-end. Among the options for a post-hoc test for non-parametric analyses, carrying out binary Mann Whitney U-tests for each of the groups is an option. That is, the first, the second and the third group will be compared with each other as 1-2, 1-3, and 2-3. The differences have been examined to see if any of the group scores are significantly different from the others. The *catch*, at this point, is the liability to Type 1 error, which is believed that there is a genuine effect in our population when, in fact, there is not. To overcome this issue, Bonferroni correction is performed (Field, 2009). This correction method is basically a restriction of the critical value to avoid Type 1 error. This is done by simply dividing the critical value (.05) by the number of the groups involved in the study. For instance, when there are three groups at hand and they are to be compared by making use of non-parametric tests, the critical value changes from .05 to .0167 (.05/3=.0167). The interpretations as to the significance of the outcomes are performed taking .0167 into account as the critical value, not the standard .05.

In the analysis process, referential and semantic comparisons were carried out among the three groups (Learner, Native 1 and Native 2). Kruskal Wallis test and Mann Whitney U-test as the post-hoc test were performed. The results of statistical analyses revealed no significant differences between learner and native texts in the anaphor reference and stem overlap indices. As for the LSA scores, bearing in mind the normality of distribution and the homogeneity of variances (Table 2), one-way ANOVA and Scheffe tests were employed; the results also revealed significant differences among the three groups.

Among the indices mentioned before, the first index from referential and semantic index set to demonstrate significant difference between L1 and L2 texts was the anaphor reference index for adjacent sentences. This index calculates the references occurring in sentences next to each other. Descriptive results for this index are provided in Table 3.

Table 3. Descriptive Results for Anaphor References for Adjacent Sentences

Groups	**n**	$\bar{x}$	**sd**
Learner (L)	49	.401	.138
Native 1 (N1)	54	.310	.142
Native 2 (N2)	46	.299	.104

Descriptive results provided in Table 3 clearly shows that the learner group (L) scored higher ($\bar{x}_L$=.401) than the other two native groups ($\bar{x}_{N1}$= .310, $\bar{x}_{N2}$= .299). To check if this difference is statistically significant, Kruskal Wallis test was conducted. The results are revealed in Table 4.

Table 4. Kruskal Wallis Test Results for Anaphor References for Adjacent Sentences

Group	n	Mean Rank	df	x^2	*p*	Group Differences
Learner	49	97.47	2	19.841	.000	L>N1&N2
Native 1	54	64.83				
Native 2	46	63				

The results of Kruskal Wallis test for adjacent anaphor reference for sentences for the three groups are displayed in Table 4. The difference among the groups appear to be statistically significant [x^2 (2) = 19.841, p< .05]. Mann Whitney U-test with Bonferroni correction as the post-hoc test reveals that this difference is between the learner and the native groups. This means that the learner group makes use of referential tools much more than the native groups regardless of the number of the words used in the texts.

The next index related to referential aspects is anaphora reference. This index takes into account the references in a given text for up to five sentences earlier. It means that it counts referential incidences backwards, be it in the first adjacent sentence or the fifth sentence backwards. Table 5 provides descriptive results for this index.

Table 5. Descriptive Results for Anaphor References

Groups	n	$\bar{x}$	sd
Learner (L)	49	.196	.086
Native 1 (N1)	54	.122	.066
Native 2 (N2)	46	.145	.056

Descriptive results for anaphor reference are displayed in Table 5. It is clear from the table that the learner group scores higher ($\bar{x}_L$=.196) than the native groups ($\bar{x}_{N1}$=.122, $\bar{x}_{N2}$=.145). In order to determine if this difference is statistically significant, Kruskal Wallis test was performed and the results are demonstrated in Table 6.

Table 6. Kruskal Wallis test results for Anaphor References

Group	n	Mean Rank	df	x^2	*p*	Group Differences
Learner	49	102.40	2	29.551	.000	L>N1&N2
Native 1	54	62.94				
Native 2	46	59.98				

The results of Kruskal Wallis test employed for the three groups in terms of anaphor reference are presented in Table 6. The analysis of the results reveals that there is a statistically significant difference among groups [x^2 (2) = 29.551, p< .05]. The results of the post-hoc test

with Bonferroni correction indicates that the there is a statistically significant difference between the learner group and the two native ones. When we refer back to the mean scores displayed in Table 5, it is obvious that the learner group scored higher from both of the native groups ($\bar{x}_L$=.196, $\bar{x}_{N1}$= .122, $\bar{x}_{N2}$= .145). This outcome indicates that in learner texts there is a plethora of references even when compared to texts, which were written by native speakers of English and which have significantly higher averages of words.

The next index in coh-metrix related to referential aspects is the argument overlap for adjacent sentences. This index calculates the overlapping arguments (nouns, verb etc.) in a given text. There are two indices which calculate argument overlaps; one adjacent overlaps and the other all overlaps across the texts. This is a proportion score and the adjacent overlap index yields ratio scores of argument overlaps between adjacent sentences. Descriptive results concerning argument overlaps between adjacent sentences are displayed in Table 7.

Table 7. Descriptive Results for Argument Overlap for Adjacent Sentences

Groups	**n**	$\bar{x}$	**sd**
Learner (L)	49	.543	.147
Native 1 (N1)	54	.562	.156
Native 2 (N2)	46	.547	.122

Table 7 reveals descriptive results for the adjacent argument overlap scores. The means of the three groups appear to be similar ($\bar{x}_L$=.543, $\bar{x}_{N1}$= .562, $\bar{x}_{N2}$=.547). Kruskal Wallis test was performed to see the statistical difference among the groups and the results as seenn in Table 8.

Table 8. Kruskal Wallis Test Results for Argument Overlap for Adjacent Sentences

Group	**n**	**Mean Rank**	**df**	x^2	p
Learner	49	71.94	2	1.003	.613
Native 1	54	79.68			
Native 2	46	72.77			

Results concerning the adjacent argument overlap scores are shown in Table 8. The analysis of the results indicate that there is statistically no significant difference among the groups [x^2 (2) = 1.003, p> .05]. This could mean that there is similar amount of adjacent argument overlaps both in the learners and the native texts. These overlaps are, in fact, related to repetitions of nouns, verbs, noun phrases, etc.; therefore, there is nothing surprising about these repetitions appearing in nearly the same amounts in adjacent sentences of all the three groups. The next index, all-distance argument overlap, tests these repetitions across the texts. Descriptive results are revealed in Table 9.

Table 9. Descriptive Results for All-distance Argument Overlap

Groups	n	$\bar{x}$	sd
Learner (L)	49	.446	.134
Native 1 (N1)	54	.449	.148
Native 2 (N2)	46	.437	.096

Descriptive results exhibited in Table 9 indicate that all-distance argument overlap scores for the three groups are quite similar ($\bar{x}_L$*=.446,* $\bar{x}_{N1}$*= .449*, $\bar{x}_{N2}$*= .437*). To verify this similarity Table 10 should be checked for the results of Kruskal Wallis test.

Table 10. Kruskal Wallis Test Results for All-distance Argument Overlap

Group	**n**	**Mean Rank**	**df**	x^2	***p***
Learner	49	73.70	2	.125	.941
Native 1	54	76.60			
Native 2	46	75.50			

Kruskal Wallis test scores concerning all-distance argument overlap for the three groups are displayed in Table 10. Again, as in the adjacent overlap scores there seems to be no significant difference among the groups [x^2 (2) = .125, p> .05]. The similarities among the groups in terms of adjacent and all-distance scores could be regarded quite normal as the subjects were given certain topics to write about and they were required to stick to them. This restriction is likely to be the cause of lexical repetitions, and thus argument overlaps appear between adjacent and distant sentences.

The next index in coh-metrix is related to stem overlaps between adjacent sentences. In this index, parts of speech aspect of lexical items enter the scene. It means that overlapping lexical items with common word roots are taken into account. For example, one sentence might include the word *lose* and the next sentence might include the word *losing* or *lost*. This incidence is counted as an adjacent stem overlap. Descriptive results concerning this index are reported in Table 11.

Table 11. Descriptive Results for Stem Overlap for Adjacent Sentences

Groups	**n**	$\bar{x}$	**sd**
Learner (L)	49	.438	.176
Native 1 (N1)	54	.560	.165
Native 2 (N2)	46	.530	.135

Descriptive results of the three groups concerning adjacent stem overlap are displayed in Table 11. According to these results, the learner group ($\bar{x}_L$*=.438*) scored particularly less than the

native ones, whereas the native group scores appear to be quite similar ($\bar{x}_{N1}$= *.560*, $\bar{x}_{N2}$= *.530*). This exclusion could be confirmed with results presented in the following table.

Table 12. Kruskal Wallis Test Results for Adjacent Stem Overlap

Group	n	Mean Rank	df	x^2	*p*	Group Differences
Learner	49	57.28	2	13.407	.001	L<N1&N2
Native 1	54	87.84				
Native 2	46	78.80				

The comparison of adjacent stem overlap scores for the three groups is presented in Table 12. The results of the comparison clearly indicate that there is a statistically significant difference among the groups [x^2 (2) = .125, p< .05]. The results of Bonferroni correction through Mann Whitney U-test makes it clear that this difference is between the learner and the native groups. This difference might be an indication of learners' lack of proficiency in modifying lexical items according to their syntactic requirements. This could also mean that learners' knowledge concerning L2 vocabulary is one-dimensional disregarding parts of speech of the lexical items at their disposal.

All-distance stem overlap is another index used in coh-metrix. In this index, stem overlaps are calculated by taking into consideration the whole text not just adjacent sentences. Descriptive results about this index are given in Table 13.

Table 13. Descriptive Results for All-distance Stem Overlap

Groups	n	x̄	sd
Learner (L)	49	.341	.155
Native 1 (N1)	54	.472	.164
Native 2 (N2)	46	.431	.110

Table 13 exhibits descriptive results for all-distance stem overlap index. Although the native groups appear to have similar mean scores ($\bar{x}_{N1}$= *.472*, $\bar{x}_{N2}$= *.431*), the learner group stands out from the native groups with a relatively low mean score ($\bar{x}_{L}$=*.341*). The following table verifies that this difference between the learner and the native groups is statistically significant.

Table 14. Kruskal Wallis Test Results for All-distance Stem Overlap

Group	n	Mean Rank	df	x^2	*p*	Group Differences
Learner	49	54.89	2	16.931	.000	L<N1&N2
Native 1	54	88.98				
Native 2	46	80.01				

Kruskal Wallis test results and binary comparison of the groups through Mann Whitney U-test with Bonferroni correction regarding all-distance stem overlap for the three groups can be checked in Table 14. The results exhibit a definite and statistically significant difference among the three groups [x^2 (2) = 16.931, p< .05]. When this difference is checked through Mann Whitney U-test to see which group was statistically excluded from the others, the scores of the learner group appears to be significantly lower than the scores of the native groups (Table 13). As was mentioned before, this index calculates the stem overlaps across a given text. Since the native groups scored significantly higher than the learner group in both adjacent and all-distance stem overlap indices, it would not be an assumption to say that the learner group lacks the ability and flexibility to make use of different parts of speech of lexical items. This index alone could be regarded as an indication of a disconnection among sentences written by the learner group.

Another index through which coh-metrix measures lexical cohesion is LSA, a statistical technique akin to factor analysis. In this study, two of these indices LSA measures for adjacent sentences and all-distance LSA measures were taken into account; and both of these indices yielded statistically significant differences among the three groups involved in the study. Descriptive results for adjacent LSA scores are detailed in Table 15.

Table 15. Descriptive Results for LSA Scores for Adjacent Sentences

Groups	**n**	$\bar{x}$	**sd**
Learner (L)	49	.185	.056
Native 1 (N1)	54	.230	.053
Native 2 (N2)	46	.233	.049

Descriptive data concerning LSA scores for adjacent sentences for the three groups are displayed in Table 15. Data revealed in Table 15 clearly indicates that the learner group has a lower mean ($\bar{x}_L$=*.185*) than both of the native groups *($\bar{x}_{N1}$= .230,$\bar{x}_{N2}$= .233)*.

As mentioned earlier, since the normality of distribution and the homogeneity of variance among LSA group scores were at acceptable levels (see Table 2), a parametric test, one-way ANOVA was employed to see if the observed mean difference among the three groups was statistically significant. One-way ANOVA and post-hoc (Scheffe) test results for adjacent LSA scores are presented in Table 16.

Table 16. One-way ANOVA Results for LSA Scores for Adjacent Sentences

	Sum of Squares	**df**	**Mean Square**	**F**	**p**	**Scheffe**
Between Groups	.072	2	.036	13.009	.000	L<N1&N2
Within Groups	.406	146	.003			
Total	.479	148				

As can be observed from Table 16, adjacent LSA scores for the groups differ significantly ($F_{(2-146)}$=13.009, p< .05). In order to determine the nature of this significant difference, a post-hoc test (Scheffe) was performed. The result of this post-hoc test clearly indicates that the learner group scored significantly lower than the native groups (L<N1&N2).

The next parameter in LSA index is all-distance scores which are calculated by taking into account LSA outcomes throughout texts. Descriptive results concerning all-distance LSA scores are presented in Table 17.

Table 17. Descriptive Results for All-distance LSA Scores

Groups	n	$\bar{x}$	sd
Learner (L)	49	.166	.05
Native 1 (N1)	54	.214	.06
Native 2 (N2)	46	.209	.05

A quick glance at Table 17 makes it clear that the learner group scored lower than the native groups ($\bar{x}_L$=.166, $\bar{x}_{N1}$= *.214*, $\bar{x}_{N2}$= *.209*). The significance of this difference is calculated by means of one-way ANOVA and Scheffe test, and the results are presented in Table 18.

Table 18. One-way ANOVA and Scheffe Test Results for All-distance LSA Scores

	Sum of Squares	df	Mean Square	F	p	Scheffe
Between Groups	.07	2	.03	11.067	.00	L<N1&N2
Within Groups	.46	146	.03			
Total	.53	148				

Analysis of all-distance LSA scores indicates that the difference among groups concerning all-distance LSA scores are statistically significant [$F_{(2-146)}$=11.067, p<.05]. Furthermore, Scheffe test result reveals that this difference statistically excludes the learner group from the native ones (L<N1&N2).

Parametric analyses of adjacent and all-distance LSA scores revealed that the learner group scored significantly low compared to both native groups. The noteworthy aspect of these outcomes is that LSA scores, both adjacent and all-distance, are not influenced by the average number of the words per text produced by the three groups. In other words, no matter what the native text lengths are, the learner group obtained significantly lower LSA scores. When the relation of LSA scores with lexical cohesion in texts is taken into consideration, written productions of the learner group can be claimed to be less cohesive compared to both of the native group written productions**.**

DISCUSSION

The results of the parametric and non-parametric comparisons of texts written by Turkish EFL learners and native speakers of English validated the ability of coh-metrix, an online database for text analysis in differentiating native and non-native written productions.

With regard to the main research question of the study, regardless of average number of words in the texts and prompts (timed or untimed; exam or free writing), there appeared to be statistically significant differences between the native and the learners' text sets. This outcome is important in that no matter how many words are used, or whatever the prompt is, the learners appear to have certain features in common in their writing, which significantly differentiate them from their native counterparts.

The significant differences among groups concerning referential and semantic indices emerged in the anaphor reference, the stem overlap and the LSA indices. All of these indices, in fact, give ideas as to the unity and cohesion in a text. In these indices, learners scored significantly different from the native groups.

The abundance of anaphor references in learners' in both adjacent sentences and all through their texts means that the learner group uses referential tools much more than both of the native groups; and this difference is available regardless of the number of the words used in the texts.

The similarity in argument overlap (nouns, verbs, noun phrases, etc.) scores of the three groups both in adjacent sentences and all across the texts is a sign of repetitions of topic-related lexical items. There seems to be no significant difference between the learner and the native groups at this point.

Group mean values concerning stem overlaps both in adjacent sentences and sentences across texts appeared to be significantly different. This difference could be interpreted as a disconnection among sentences written by the learner group. In addition, learner groups' vocabulary could be regarded as dimensionally limited, as stem overlap index takes into account different parts of speech of lexical items in a given text.

LSA scores also yielded statistically significant results among the three groups; the learner group scores significantly differed from those of the native scores. Apparently, the learner group can not sustain lexical cohesion in their written productions, which is most likely to stem from the lack of lexical proficiency and flexibility in the target language.

The outcomes discussed in the previous section are all in line with the related literature. LSA related outcomes confirm Silva's (1993) findings highlighting weak lexical and semantic ties in learners' written productions. These outcomes also support the findings mentioned in Crossley et. al. (2009) that, in a text, LSA scores are strong indications of lexical relatedness.

Hinkel's (2002) findings stating that even at advanced levels EFL learners have severely limited lexical and syntactic repertoire, are also confirmed in terms of lexical repertoire. The significant differences between the learner and the native groups in terms of stem overlap both in adjacent sentences and all across texts could be counted as an outcome of this severe limitation.

CONCLUSIONS

This study validates an online database used to evaluate texts, both native and learner, at multiple levels. Some indices in this database managed to differentiate between native and learner text sets. Moreover, the same indices were unable distinguish two different native corpora. From a language learning-acquisition point of view, this is important in that it paves the way to the possibility of grading learners' texts digitally, which would solve the problem of subjectivity in language testing and evaluation process.

The study also revealed some problematic areas in learners' texts while some other aspects were reconfirmed. However, without checking it with our subjects' native language writing skills, it is hard to say that these outcomes are universally valid for language learners. That is to say, our subjects might already be *unskilled writers* in their native language trying to survive a foreign language by making do with whatever linguistic repertoire they have at their disposal.

REFERENCES

Allen, D. (1992). *Oxford placement test 2 (New edition).* Oxford: Oxford University Press.

Brown, D. H. (2007). Principles *of language learning and teaching*. New York: Longman.

Connor, U. (1984). A Study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics* 17 (1-4), 301-316.

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

Crossley, S. A., Salsbury, T. McCarthy, P. M. & McNamara, D. S. (2008). Using Latent Semantic Analysis to explore second language lexical development. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the 21st International Florida Artificial Intelligence Research Society* (pp. 136-141). Menlo Park, California: AAAI Press.

Crossley, S. A. & McNamara, D. S. (2009). Computationally assessing lexical differences in second language writing. *Journal of Second Language Writing*, 17 (2), 119-135.

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, 16, 409–435.

Ellis, R. & Barkhuizen G. (2005). *Analysing Learner Language* (pp. 1-15). Oxford: Oxford University Press.

Ferris, D.R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly* 28, 414-420.

Field, A. (2009). *Discovering Statistics Using SPSS*. London: Sage Publications.

Flowerdew, J. (2009). Use of signalling nouns in a learner corpus. In Mahlberg & Flowerdew. *Lexical Cohesion and Corpus Linguistics*. Amsterdam: John Benjamins.

Frase, L., Faletti, J., Ginther, A., & Grant, L. (1999). *Computer analysis of the TOEFL Test of Written English* (Research report 64). Princeton, NJ: Educational Testing Service.

Gass, S., Selinker, L. (2008). *Second Language Acquisition: An Introductory Course*. New York, NY: Routledge.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193–202.

Granger, S. (1993). The International Corpus of Learner English. In Aarts J., de Haan P. and Oostdijk N. (Eds.) *English Language Corpora: Design, Analysis and Exploitation*. (pp. 57-69). Amsterdam: Rodopi.

Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly* 37(3), 538-545.

Hinkel, E. (2002). *Second Language Writers' Text*. Mawah, NJ: Lawrence Erlbaum Associates.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.

Meara, P. (2002). The discovery of vocabulary. *Second Language Research*, 18(4), 393-407.

Pica, T. (2005). Second Language Acquisition Research and Applied Linguistics. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp.263-280). Mahwah, N.J: Lawrence Erlbaum Associates.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657-77.