A Computational Investigation of Cohesion and Lexical Network Density in L2 Writing

Clarence Green[1]

[1] School of Languages and Linguistics, University of Melbourne, Melbourne, Australia

Correspondence: Clarence Green, Bld. 139 Parkville, Melbourne 3010, VIC, Australia. Tel: 1-03-8344-8032. E-mail: c.green4@pgrad.unimelb.edu.au

Abstract

This study used a new computational linguistics tool, the Coh-Metrix, to investigate and measure the differences in cohesion and lexical network density between native speaker and non-native speaker writing, as well as to investigate L2 proficiency level differences in cohesion and lexical network density. This study analyzed data from three corpora with the Coh-Metrix: the International Corpus of Learner English (ICLE) as an L2 higher proficiency group, the Louvain Corpus of Native English Essays (LOCNESS) as a native speaker baseline, and a collected EFL corpus from Indonesia for the L2 lower proficiency data.

Statistical investigation of the Coh-Metrix results revealed that five out of six Coh-Metrix variables used in this study did not detect proficiency level differences in L2 but the tool was consistently able to distinguish between L2 and native speaker writing. Differences included that L2 writing contains more argument overlap, more semantic overlap, more frequent content words, fewer abstract verb hyponyms and less causal content than native speaker writing.

Keywords: cohesion, NLP, second language writing, corpus linguistics, computational linguistics

1. Introduction

Learning to write extended discourse in a second language is a difficult skill for second language learners. It is also a fundamentally important skill for many non-native speakers who need to develop a command of written English for academic and professional success (Silva, 1993). Language mechanics such as orthography, punctuation and lexical selection have long been established areas of difficulty for L2 writers, and salient features which mark L2 writing as non-native like (White, 1987). Regarding advanced academic prose, Cumming (2001), in a review of twenty years of empirical studies on second language writing, identified that the most difficult developmental areas include the complex syntax, rhetorical strategies and specificity of vocabulary needed for the academic register.

For teachers and assessors, L1-L2 differences in language mechanics are relatively easy to objectively discern and measure (Bardovi-Harlig & Bofman, 1989). It is more difficult, however, to investigate and quantitatively measure L1-L2 differences in textual cohesion and lexical network density, and it unclear how these develop in non-native speakers as proficiency increases. These areas need research attention. Cohesion is a crucial skill that L2 writers need for academic success (Mirzapour & Ahmadi, 2011), as without the ability to create cohesion through the appropriate use of language, texts are rendered difficult to follow (Halliday & Hasan, 1976). Being able to define how native and non-native speakers differ with regard to cohesion and lexical network density, as well as knowing how these features differ across L2 proficiency levels, would be beneficial for understanding L2 writing development, for designing instruction, and for validating writing tests.

1.1 Computational Tools

Computational tools such as ETS's eRater (Attali & Burstein, 2006) and other corpus linguistics' software are able to investigate L2 lexical differences and mechanics, however until recently no computational system has been able specifically analyse cohesion and lexical network density. Advances in computational linguistics and natural language processing (NLP) have made available a comprehensive new software tool, the Coh-Metrix, which has the potential for a deep level quantitative investigation of textual cohesion and lexical network density in second language writing. The system draws together research from a variety of disciplines including discourse

analysis, psycholinguistics, corpus linguistics and natural language processing, making use of previous computational systems by incorporating WordNet (Miller at al, 1995), the CELEX database (Coltheart, 1981), the MRC psycholinguistics database (Baayen, Piepenbrock, & Gulikers, 1995), Latent Semantic Analysis, as well as a range of other part of speech taggers, lexicons, and semantic interpreters.

Crossley and McNamara (2009) first used the Coh-Metrix to measure the differences in cohesion and lexical network density between native speaker and non-native speaker writing. They concluded that the Coh-Metrix demonstrated that L2 writers used less explicit cohesive markers and had less dense lexical networks than native speakers. However, these conclusions were based on a Coh-Metrix analysis of data with a single L1 background. The current research aimed to test the validity of their claims that their findings were general L1-L2 differences. This was done using a range of L1 backgrounds from the International Corpus of Learner English (ICLE) in a methodologically comparable Coh-Metrix study. The current research also aimed to establish whether the Coh-Metrix could be used to investigate the development of L2 lexical network density and textual cohesion across L2 proficiency levels toward a native speaker standard.

1.2 Textual Cohesion and L2 Writing Development

Cohesion consists of related lexical and grammatical markers throughout discourse to facilitate coherence, and is a means by which speakers meet communicative goals effectively (Schiffrin, 1987; Witte & Faigly, 1981). Learners of English need to acquire the ability to manage cohesion to achieve communicative competence (Cumming, 2001; Mirzapiur & Ahmadi, 2011).

Halliday and Hasan (1976) describe five core classes of cohesion: substitution (e.g. one, any), ellipsis, reference cohesion (e.g. pronouns), conjunctive cohesion (e.g. coordinators, adverbials) and lexical cohesion (e.g. repetition, synonymy). These classifications have provided the framework for most research into the relationship between textual cohesion and L2 proficiency, and for research into the differences between cohesion in L1 and L2 writing. However, both research programs have produced complex and non-definitive results (Chen, 2008; Granger & Tyson, 1996; Silva, 1993). Previous studies constitute three broad categories: studies that have shown a positive correlation between the number of cohesive devices and proficiency level, studies that have shown no significant relationship, and studies that have shown an inverse relationship where more cohesive devices in an L2 text correlate with lower proficiency levels.

Ferris (1994) found a positive correlation between cohesion and L2 proficiency in a study of 160 Arabic, Chinese, Japanese and Spanish ESL writers. Regardless of L1 background, Ferris' (1994) study showed that writers at a higher proficiency used more of the cohesive devices of repetition, synonymy and pronominal reference than the lower proficiency writers. Liu and Braine (2005), using essay data from 50 undergraduate Chinese EFL writers, also found a strong positive correlation between writing ability and the prevalence of cohesive devices. Reference chains consisting of anaphoric pronominals, demonstratives, articles and comparatives correlated very highly (r =.851, p < .05) with high proficiency essays. Cohesion through lexical repetition, synonymy, collocation and hyponymy also correlated strongly (r =.842, p < .05) with proficiency. Positive correlations between cohesion and proficiency were also found by Field and Yip (1992), Norment (2002) and Faigly (1981).

Castro (2004), conversely, found an insignificant relationship between cohesive devices and L2 proficiency in an investigation of EFL data from a Philippines' university. Using low, intermediate and high proficiency data, she found that the cohesion through pronominals, articles, demonstratives and comparatives showed no statistically significant differences across proficiency levels. Lexical cohesion, as with Liu and Braine (2005) also exhibited no significant differences, nor did markers of conjunctive cohesion significantly differ across proficiency levels. Further studies have shown an inverse relationship between frequencies of cohesive devices and L2 writing proficiency. Crossley and McNamara (2011) analyzed 1200 EFL essays from Hong Kong High School graduation exams and found that the more advanced writers used fewer cohesive devices. Focusing on cohesion through aspect repetition, lexical word frequency, meaningfulness and familiarity, Crossley and McNamara (2011) found a 'reverse cohesion effect' where more advanced writers used fewer cohesive devices being more confident of reaching their communicative goals. One might hypothesize that as L2 writing proficiency increases and a better understanding of how the target language can be used to reach communicative goals develops, learners would become more economical in their use of cohesive devices.

Turning to research that has directly compared textual cohesion between L1 and L2 writing, a picture emerges as complicated as that of cohesion and proficiency level. Mirzapour and Ahmadi (2011) compared 60 English and Persian research articles and found pattern differences in the distribution of lexical cohesion with Persian writers having a general tendency to use lexical repetition and collocations. Ferris (1994b) compared 30 NS and 30 ESL

persuasive writing samples and found that the ESL group used significantly more cohesive adjuncts (e.g. however, firstly, secondly, in conclusion) than native speakers. Kenkel and Yates (2009) also found an overuse by L2 writers of noun phrase repetition and pronominal reference between sentences. On the other hand, Crossley and McNamara (2009), using L2 persuasive essays from Spanish L1 writers, found less cohesion than in their native speaker data. Finally, Johnson (1992) compared 20 Malay ESL and 20 NS persuasive essays and found that there were no statistically significant difference between L1 and L2 cohesion.

The complex and often contradictory research findings can be partially explained by the fact that studies often define the construct of cohesion differently. For example, Green, Christopher, and Mei (2000) investigated textual cohesion qualitatively in Chinese EFL writing through topic fronting and logical connectors, whereas Chen (2008), using quantitative corpus methodology investigated Chinese EFL writing through anaphors, lexical overlap and conjunctive devices. Chen (2008) claims there is no relationship between highly cohesive texts and EFL proficiency, whereas Green, Christopher, and Mei (2000) concludes that greater use of cohesive devices correlates with poor writing.

The computational accuracy and reliability of the Coh-Metrix may help to clarify the complex and often contradictory findings which are particularly problematic in L2 cohesion research.

1.3 The Development of Lexical Network Density and L2 Vocabulary

Related to the ability to maintain cohesion is the extent of speakers' lexical networks- the complex of associations connected with the semantic properties of a lexeme (Hudson, 2008). The network not only includes a speaker's knowledge of the semantic relationships of a lexeme, its hypernyms/hyponyms, synonyms, antonyms and so on, but also its morphosyntactic and phonological properties (Lyons, 1968). The density of a lexical network is the number of (non-random) connections a network contains, which increases during L2 (and L1) development as more associations are incorporated. Empirically, however, the lexical network has been an under-studied subfield of SLA (Crossley & McNamara, 2009). However, exceptions include Schmidt and Meara (1997) who tested 95 Japanese EFL students on word association tests in a 12 month longitudinal study, showing that L2 vocabulary gain scores correlated significantly with the semantic connections EFL learners were able to make.

Meara (2007) examined L2 lexical networks compared to native speakers. He compared English learners of French and French NSs on multiple tasks in which participants selected 2 French words (out of 5) they felt to be connected. Results showed that as L2 proficiency developed so did lexical network density since associations made by the English learners became more native like as proficiency increased.

1.4 Coh-Metrix: The Analysis of Cohesion and Lexical Network Density

The Coh-Metrix is a computational linguistics tool that was designed for the analysis of cohesion in native speaker written discourse so that text readability could be matched to educational levels, ensuring developmentally appropriate material for the education of native English speaking (NES) students (Graesser et al, 2004). However, the Coh-Metrix was not designed as an L2 tool, and Crossley and McNamara (2009) were amongst the first to employ it for L2 cohesion analysis. They also extended the function of the tool to the analysis of L1/L2 lexical network density on the basis that the Coh-Metrix includes measures of semantic associations, hypernymy, synonymy, polysemy and other variables which need not be taken as cohesion specific.

The tool has undergone multiple validations (and version updates) to confirm that it measures the cohesion construct (but not as yet the lexical network density construct), and has been made available online by the University of Memphis (http://cohmetrix.memphis.edu). The tool incorporates previous advances in computational linguistics such as WordNet (Miller at al, 1995), the MRC Psycholinguistics Database (Coltheart, 1981), and the CELEX Database (Baayen, Piepenbrock, & Gulikers, 1995). Together, these resources allow the Coh-Metrix to process natural language and describe features such as semantic associations, hypernymy, part of speech type/token and word frequency, along with dozens of other variables not used in the current study.

Crossley, Greenfield and McNamara (2008) first employed the Coh-Metrix as a readability tool that might be employed to design TESOL teaching material. A corpus of short written passages were evaluated for readability by Japanese EFL learners and then correlated with traditional readability measurements, including the Flesch Reading Ease Scale. The corpus was then analysed for cohesion by the Coh-Metrix and the results correlated more strongly with student perceptions of readability than the traditional readability formulas. This indicated the Coh-Metrix could help determine teaching materials which would be appropriate for an L2 audience.

The current study builds upon the foundations of Crossley and McNamara (2009) in which, through a Coh-Metrix analysis of ICLE data, 10 variables out of the many available in the tool showed the most significant

59

differences between L1/L2 writing. They processed 195 Spanish advanced proficiency EFL persuasive essays and 208 NS persuasive essays from the ICLE through the Coh-Metrix. Statistical analysis identified that out of all the indices available in the Coh-Metrix, measurements on 10 variables most strongly indicated differences between the L1 and L2 data. A further discriminate function analysis showed that measurements on 7 of these variables were capable of predicting essay language background with 79% accuracy. The variables were grouped by Crossley and McNamara (2009) into measures of cohesion and measures of lexical network density, with the study concluding with the strong claim that L2 writing, in general, is characterized by less cohesion than L1 writing and that L2 learners have significantly less density in their lexical networks. This is a strong conclusion because the study had only one L1 background represented in their L2 data, which may not reflect patterns of cohesion in L2 writing in general, and because there may be a problem with generalising properties of written data at a particular proficiency level to the psycholinguistic realities of L2 lexical networks (i.e. should the semantic associations made in a text be taken as a direct reflection of semantic knowledge). It was felt that a follow up study was needed which used the same variables and similar methodology to Crossley and McNamara (2009) but incorporated different proficiency levels and a range of L1 backgrounds.

2. This Study

Following Crossley and McNamara's (2009) study, it was of interest to see whether 6 of the 10 Coh-Metrix variables (not all 10 variables were available on the publicly accessible version of the Coh-Metrix used in this study) shown by their study to best indicate L1/L2 differences would show development across sequential proficiency levels in their lexical network density and cohesion as learners approached the target language. The results of this investigation were felt to be indicative of the full Coh-Metrix's prospects as a tool for tracking L2 development.

This study also wanted to validate Crossley and McNamara's (2009) general claims about features of L2 writing based their results on the 6 Coh-Metrix variables. They claimed that the tendencies shown by the Coh-Metrix indicated that compared to native speaker writing, L2 writing is marked by less causal content, less noun repetition, fewer pronoun references, less semantic connections across sentences, less hypernymic abstraction, and a higher proportion of frequently occurring content words. Based on these features, Crossley and McNamara (2009) concluded that L2 writers do not provide as many cohesive devices in their text as native speakers, and that they have less dense lexical networks. However, their research claims were based on the analysis of a single L1 background (the ICLE advanced proficiency Spanish L1 subcopora). In this study, ICLE data from a variety of L1 backgrounds was used. This would determine whether Crossley and McNamara's (2009) findings were indeed generalizable features of L1/L2 writing difference as they concluded, or were rather a product of their study's controlled L1 background and therefore reflective only of their data.

3. Research Questions

Two research questions were specifically addressed:

1: In the 6 key Coh-Metrix measurements of cohesion and lexical network density that were established by Crossley and McNamara (2009) to strongly indicate L1/L2 writing difference, can the Coh-Metrix illustrate progression on these variables across lower and higher EFL proficiency levels toward NS norms?

2: Will a Coh-Metrix analysis of ESL written data from a variety of L1 backgrounds, and different proficiency levels, confirm less cohesion and lexical network density are general features of L2 writing?

4. Methodology

4.1 Data

The data for this study consisted of persuasive essay writing samples selected from four different sources: the International Corpus of Learner English (henceforth ICLE) (Granger et al, 2009), the Louvain Corpus of Native English Essays (henceforth LOCNESS) (Granger, 1995), and two corpora collected from separate EFL schools in Indonesia. Topics were varied, however, two persuasive essay topics were common and appeared across all corpora: the importance of environmental conservation and the pros and cons of nuclear power. None of the persuasive essays were timed, or produced under test conditions.

The ICLE is a corpus of learner English and provided the higher proficiency level L2 data for this study. The corpus consists of EFL essays, averaging 617 words, from 16 different L1 backgrounds. Male and female students are represented, and average participant age is 22. Each L1 backgrounds forms a subcorpora consisting of approx. 200,000 words of learner writing collected from university undergraduates while attending university in their native countries. The proficiency level in the corpus consists of advanced learner data from English majors in their third or fourth year of university. The ICLE data having been independently rated for advanced

proficiency level based on the Common European Framework of Reference for Languages (CEF) (Granger et al, 2009). The majority (12/16) of first languages represented are Indo-European, the exceptions being Japanese, Chinese, Turkish and Setswana. The ICLE was used in this study for comparability, as it was the corpus used by Crossley and McNamara's (2009), in which the Spanish L1 subcorpus was used.

The native speaker baseline data for the study was taken from the LOCNESS corpus. This corpus consists of 149,574 words of persuasive essays above 500 words written by American university students, a further 59,568 words of persuasive and literary essays written by British university students, and 60,209 words of British A-level (pre-university) persuasive essays. Both genders are represented, and the age range is from 18-21. The LOCNESS was selected as it was the NS corpus used by Granger and Tyson (1996), whose research supported a correlation between the overuse of cohesive devices in L2 writing when compared to L1 writing, operationalized through the LOCNESS.

The lower proficiency L2 data consisted of two intermediate level corpora collected from Indonesian EFL Schools. Given that the ICLE was an EFL corpus, for comparability EFL data for the intermediate proficiency level was used. A corpus of persuasive essays was collected from the Indonesia Australia Language Foundation (IALF). The IALF is an English language training school in Surabaya, Indonesia. The school is an IELTS test centre for pre-university students, specializing in English for Academic Purposes (EAP). The majority of students who graduate the IALF enrol in Australian undergraduate courses. The data consisted of essays that had been collected by teachers over previous years for auditing purposes. Most students were between 14-17 years of age. The IALF corpus was felt to match well the ICLE as it consisted of academic writing in a persuasive form and many of the essay topics matched both the ICLE and LOCNESS.

As only one L1 background was represented in the IALF corpus, to complete the lower proficiency data group another corpus containing a variety of L1 backgrounds was collected from an international high school in Jakarta, Indonesia. This institution differed from the IALF in that it was not primarily a language training centre, but a high school for international students where content instruction was entirely in English. The aims of the school regarding English are identical to the IALF as English language teaching focused on academic English to enable students to be accepted into undergraduate courses at western universities. The EFL essays from the international school had a variety of L1 backgrounds including Chinese, Javanese, Malaysian, Korean and Indonesian, though the exact numbers for each were unknown. The age range was from 15-17. The mean essay length for the Indonesian data was 367 words (Table 1, section 4.3).

While the data from the ICLE had been vetted as being advanced and university level by the corpus designers, and the Indonesian data derived from pre-university and high school learners, nonetheless previous exposure to English instruction for the two groups was unknown and further confirmation that the Indonesian corpus was indeed of a lower proficiency level than the ICLE was needed. To establish that the Indonesian data truly represented a lower proficiency level, a random sample was taken from both data and independently rated by two raters. The first rater was an experienced academic with EFL teaching experience in Germany, and the second rater had masters' level qualifications in TESOL. Every third essay from both corpora (n=28, 22% of the overall data used in this study, see section 4.3) was given to the raters to holistically assign to a high or low proficiency group based on their teaching experience. All ICLE data was assigned to the high proficiency group, and all Indonesian data was assigned to the lower group, with 100% agreement between the raters. This suggested that there were clear proficiency level differences in the two corpora.

4.2 Instruments

Corpus data was processed with the Coh-Metrix, a web-based NLP tool with over 50 variable measurements of cohesion and lexical network density. In Crossley and McNamara (2009), measurements on 10 of these variables showed the greatest differences between L1/L2 writing. In this study, 6 of these variables were used to analyse the three corpora. The following provides a brief description, according to the Coh-Metrix developers (Granger et al, 2004; McNamara et al. 2006), of how each of the variables measures an aspect of the cohesion and lexical network density constructs. The findings of Crossley and McNamara (2009) on each of the variables are also specified to facilitate later comparison with this study's results:

4.2.1 Variable 1: Causal Content

This variable measures the extent to which cohesion is signalled in a text through cause and effect relationships. More cause and effect relationships in a text are argued to be conducive to cohesion as one element is marked as leading to another (Granger et al, 2004; Halliday & Hasan, 1976). The Coh-Metrix measures this through a frequency count of the amount of causal verbs and causal particles in a text. Causal particles consist of linking content across clauses, including items such as *since*, *so that*, *because*, and *consequently*. Causal particles are

identical to the causal devices in the conjunctive cohesion category of the Halliday and Hasan (1976) framework (section 1.2).

Causal verbs, the other item contributing to the causal content variable, are verbs which have been classified by WordNet as semantically indicating the cause of a change of state. For example, *kill* is tagged by WordNet as causal because it signifies 'cause to die'. Crossley and McNamara (2009) found that the incidence of causal verbs in L2 writing was significantly less than in L1 writing.

4.2.2 Variable 2: Adjacent Argument Overlap

Adjacent argument overlap is a cohesion measure for which the Coh-Metrix provides a cosine value (0-1) reflecting the level to which adjacent sentences share one or more arguments. Arguments are defined as reference chains consisting of a noun or noun phrase which is either repeated or used as the antecedent for later pronoun reference. A figure provided by the Coh-Metrix as approaching 1 represents a highly cohesive text. Crossley and McNamara (2009) found a low argument overlap was distinctive of L2 writing in their study.

4.2.3 Variable 3: Latent Semantic Analysis (LSA) Sentence Adjacent

LSA is an NLP statistical procedure that evaluates the shared semantic information between adjacent sentences by calculating the amount of new information and semantically given information across all lexical words. The following example from the L2 lower proficiency data group illustrates a high amount of shared semantic content between sentences:

According to me, the best things living in a big <u>city</u> is <u>fun</u> and many <u>places</u> that we can <u>visit</u>. For example, <u>mall</u>, <u>theatre</u>, and other public <u>places</u>, like <u>railway station</u>, <u>harbour</u>, <u>airport</u>, <u>zoo</u>, etc. If we feel <u>bored</u> and need refreshing, we can <u>go</u> to one of the <u>place</u> above.

This text receives a high LSA score because of its many semantic associations. It is semantically 'given' that *mall*, *theatre, railway station* and so on entails that they are *places*. The Coh-Metrix calculates the LSA cosine value from 0 to 1. Values approaching 1 signal a large amount semantic overlap. LSA is sensitive to 'Lexical Chains' or 'Lexical Sets' (Collins and Hollo, 2010), and collocations shown to be definitive features of L2 lexical cohesion by Mirzapour and Ahmadi (2011). Crossley and McNamara (2009) found their L2 data had lower LSA values than their NS corpus, and they claimed this signalled less dense lexical networks.

4.2.4 Variable 4 and 5: Noun and Verb Hyperrnymy

A hyperrnym is a superordinate term with the semantic field of other words falling within it. Hypernymy measures indicate the strength of lexical networks (Crossley & McNamara, 2009) because hypernymic relations reflect a speaker's lexical knowledge of specific and abstract words within the same semantic hierarchy. Knowing which lexical items contain the semantic content of others is part of having a well developed linguistic system. (Cruse, 1986; Lyons, 1995). The Coh-Metrix provides a mean value from 0-7 for each text averaged from the hypernym hierarchy position of every verb and noun in that text according to the WordNet database. A Coh-Metrix mean closer to 0 represents a text that contains mostly specific words, whereas a mean closer to 7 represents a text with many highly abstract words.

4.2.5 Variable 6: Frequency of Content Words (CELEX Written Frequency)

This variable draws on the CELEX word frequency database. The Coh-Metrix ranks all content words in a text according to their frequency of occurrence in the English language. Through a logarithm, this variable provides a mean from 0-6 for each text, with higher values reflecting that a text contains more words of frequent occurrence.

4.3 Procedure and Analysis

Corpus data was grouped in three categories, L2 Higher Proficiency, L2 Lower Proficiency and L1 Native Speaker. To establish the L2 High Proficiency, 50 persuasive essays were taken from the ICLE, 10 each from different language backgrounds. All essays selected were between 250-600 words. The L1 backgrounds selected were German, Russian, Japanese, Turkish and Setswana. These languages were selected in order to avoid the overrepresentation of one particular language family in the group. This was important as, for example, if three Germanic languages and 2 Romance languages comprised the high proficiency group, the study's answers to Research Question 2, regarding generalizable indicators of L2 writing difference, would have been called into question. Thus the High Proficiency group consisted of a Romance language, a Germanic language, a Slavic language, an Asian language, a Bantu language and an Altaic language.

The L2 Lower Proficiency group consisted of 34 persuasive essays from the Indonesian corpora. From the IALF corpus 23 essays were selected with a further 11 essays taken from the Jakarta international school corpus. All

L2 Lower Proficiency essays selected were between 250-550 words. The L1 Native Speaker data group consisted of the Marquette University essay subcorpus of the LOCNESS. This comprised 46 NS persuasive essays from an American tertiary institution. These essays were significantly longer than either the ICLE or Indonesian essays, so only the first five paragraphs of each essay (if the fifth paragraph exceeded 600 words, the fourth paragraph was made the cut off point) were used, providing word counts ranging from 250-600 words. There were no grounds to believe that complete essays were required for the variables on the Coh-Metrix to be efficacious since there is no reason why cohesive devices would be skewed in their distribution throughout a text.

Table 1. Average Word Counts and Analysis of Variance from the three data groups

Data Group	N	Mean	Std. Deviation	F-value
L2 Lower proficiency	34	367	95	2.257 (n.s.)
L2 Higher proficiency	50	409	120	
L1 Native Speaker	45	411	81	

Once the three data groups had been established, all texts (N=129) were processed through the Coh-Metrix. The texts, which were Microsoft Word files, were individually cut and pasted into the Coh-Metrix main window and submitted for analysis. The Coh-Metrix software returned a numerical output for each essay on the 6 variables (full description of calculations at http//:cohmetrix.memphis.edu). This numerical output was then analyzed statistically using SPSS v 17 with the level for significance set at $p < .05$. To answer Research Question 1, a series of one way ANOVAs were run to compare the data groups cohesion and lexical network density on each variable.

5. Results

5.1 RQ 1: Will the Coh-Metrix Illustrate Progression across a Low and High EFL Writing Proficiency Level toward NS Norms?

5.1.1 Variable 1: Causal Content (Scale 1- ∞)

As shown in Table 2, the ANOVA for this variable showed significant differences between the groups ($F(2, 126) = 9.50$, $p < .05$, $\eta^2p = .131$) although the effect size was relatively small.

Table 2. Variable 1: Causal Content, Analysis of Variance from the three data groups

Data Group	N	Mean	Std. Deviation	F-value
L2 Lower proficiency	34	71.26	18.52	9.50
L2 Higher proficiency	50	62.93	14.17	p< .05
L1 Native Speaker	45	76.75	16.60	

A Levene's test indicated homogeneity of variance, so a post hoc Scheffe test (Table 3) was run to establish which groups were significantly different from one another.

Table 3. Variable 1: Causal Content, Comparison between the three data groups

Data Group	Comparison Group	Mean Difference	Significance level
L2 Lower Proficiency	L2 Higher proficiency	8.53	.052 (n.s.)
L2 Lower Proficiency	L1 Native Speaker	n.s.	n.s.
L2 Higher Proficiency	L1 Native Speaker	13.82	.001

A significant difference was shown to exist between the amount of causal content in texts written by the L2 Higher Proficiency group and those by L1 Native Speakers, but not between the native speaker texts and the L2 Lower Proficiency. Closely approaching significance (p=.052) was the difference between the L2 proficiency

levels. The analysis indicates that the use of causal content as a cohesive device does not follow a linear progression in which lexical items signalling causation increases as L2 proficiency develops. Rather, a U-shaped developmental sequence is shown.

5.1.2 Variable 2: Argument Overlap (Scale 0-1)

A statistically significant ANOVA was found for the overlap between sentence adjacent arguments ($F(2, 126)$ = 131.92, $p < .05$, $\eta^2 p = .677$) with a large effect size.

Table 4. Variable 2: Argument Overlap, Analysis of Variance from the three data groups

Data Group	N	Mean	Std. Deviation	F-value
L2 Lower proficiency	34	.610	.167	131.92
L2 Higher proficiency	50	.600	.210	$p < .05$
L1 Native Speaker	45	.121	.067	

Unequal variances were signified by a Levine's test, so a post hoc Tamhane comparison of means (Table 5) was run to identify differences amongst the groups.

Table 5. Variable 2: Argument Overlap, Comparison between the three data groups

Data Group	Comparison Group	Mean Difference	Significance level
L2 Lower Proficiency	L2 Higher proficiency	n.s	n.s.
L2 Lower Proficiency	L1 Native Speaker	.490	.001
L2 Higher Proficiency	L1 Native Speaker	.481	.001

Both L2 data groups used significantly more argument overlap between sentences than native speakers (i.e. more noun phrase repetition and extended anaphoric reference chains), though they did not differ significantly from each other. No linear progression is evident on this variable. The difference between the L1/L2 groups is striking, with L2 writing containing approximately 5 times more argument overlap than native speakers (based on mean estimates).

5.1.3 Variable 3: Latent Semantic Analysis (Scale 0-1)

The ANOVA for this variable showed that the amount of lexical connections in texts, judged through the amount of 'given' semantic information across sentences, was significantly different amongst the groups ($F(2, 126)$ = 44.397, $p < .05$, $\eta^2 p = .413$) with a medium effect size.

Table 6. Variable 3: LSA sentence adjacent, Analysis of Variance from the three data groups

Data Group	N	Mean	Std. Deviation	F-value
L2 Lower proficiency	34	.247	.081	44.397
L2 Higher proficiency	50	.202	.070	$p < .05$
L1 Native Speaker	45	.117	.031	

To indicate the differences, Tamhane's post hoc comparison of means was appropriate (Table 7) as a Levene's test indicated unequal variances.

Table 7. Variable 3: LSA sentence adjacent, Comparison between the three data groups

Data Group	Comparison Group	Mean Difference	Significance level
L2 Lower Proficiency	L2 Higher proficiency	.045	.05
L2 Lower Proficiency	L1 Native Speaker	.131	.001
L2 Higher Proficiency	L1 Native Speaker	.085	.001

All groups were significantly different from each other. A progressive linear decline in the amount of given semantic information across sentences is evident as proficiency increased toward native speaker norms.

5.1.4 Variable 4: Noun Hypernymy (Scale 0-7)

There were no significant differences in the level of noun hypernymy across the 3 groups (F(2, 126) = 2.185, n.s). A helpful anonymous reviewer suggested there must be reasons that would account for the lack of significance between native speakers, and lower and higher L2 noun hypernymy, particularly since verb hypernymy (table 8) had differences interpretable as a development from concrete to abstract as proficiency increased. This researcher unfortunately cannot draw from this result alone a principled explanation which he feels has much strength beyond speculation. Indeed, the result must not mean that there are no differences in the knowledge and level of abstractions of nouns between native speakers and L2 learners, as there undoubtedly must be. Possibly, a stylistic/genre effect is at work on nouns in the data, given they were in persuasive essays on generalist topics. Generalist persuasive essays perhaps need not require a great deal of nominal abstraction, leading to an assessment of the data groups as similar by the Coh-Metrix. For example, more specialist topics may have brought out more abstract nominals and highlighted limitations in L2 vocabulary. Of course, one would also need to conclude this genre effect was restricted to nominals as it did not affect verb abstraction. It may be that knowledge of verb hypernymy develops faster than nouns, and/or more abstract verbs are used more frequently than more abstract nouns.

5.1.5 Variable 5: Verb Hypernymy (Scale 0-7)

Unlike the levels of noun hypernymy, there were significant differences in verb abstraction according to mean hypernym levels across the data groups (F(2, 126) = 17.16, p< .05, η^2p = .214), though the effect size was small.

Table 8. Variable 5: Verb hypernym levels, Analysis of Variance from the three data groups

Data Group	N	Mean	Std. Deviation	F-value
L2 Lower proficiency	34	1.44	.025	17.16
L2 Higher proficiency	50	1.49	.023	p< .05
L1 Native Speaker	45	1.63	.020	

The means in Table 8 show that all groups had verbs with rather low hypernymy values (the scale being 0-7). Therefore, there was a preference across all groups for concrete verbs as opposed to higher level hypernyms which are more abstract. To identify what contributed to the ANOVA's significance, a post hoc Scheffe test was run (Table 9).

Table 9. Variable 5: Verb hypernym levels, Comparison between the three data groups

Data Group	Comparison Group	Mean Difference	Significance level
L2 Lower Proficiency	L2 Higher proficiency	n.s.	n.s
L2 Lower Proficiency	L1 Native Speaker	.188	.001
L2 Higher Proficiency	L1 Native Speaker	.133	.001

The L2 proficiency levels did not differ significantly in their verb hypernym levels, indicating that both similarly use verbs with concrete and specific meanings rather than more abstract superordinate terms. Native speaker's verbs, however, contained a higher level of hypernymic abstraction. While the means demonstrate a linear growth in hypernymic abstractions across proficiency level toward the NS norm, the L2 differences were not statistically significant and as such one cannot conclude progressive development, and that verbs become more abstract as proficiency increases.

5.1.6 Variable 6: Frequency of Content Words (Scale 0-6)

The prevalence in the data of the most frequent content words of English showed a significant difference (F(2, 126) = 31.458, p< .05, η^2p = .352) amongst the groups (Table 10), with a medium effect size.

Table 10. Variable 6: Frequency of content words, Analysis of Variance from the three data groups

Data Group	N	Mean	Std. Deviation	F-value
L2 Lower proficiency	34	2.43	.127	31.458
L2 Higher proficiency	50	2.50	.162	p< .05
L1 Native Speaker	45	2.23	.183	

The data groups were shown to have an equality of variances, so a post hoc Scheffe test was run (Table 11) to indicate the source of the significant ANOVA.

Table 11. Variable 6: Log. Frequency of content words, Comparison between the three data groups

Data Group	Comparison Group	Mean Difference	Significance level
L2 Lower Proficiency	L2 Higher proficiency	n.s.	n.s
L2 Lower Proficiency	L1 Native Speaker	.120	.001
L2 Higher Proficiency	L1 Native Speaker	.268	.001

The L2 Higher Proficiency group differed from the native speakers in using more words which have a higher frequency of occurrence in English. The L2 Lower Proficiency group also used more frequently occurring content words than the native speakers, but they did not use a statistically different amount than the Higher Proficiency group. As with Variable 7 (verb hypernym), and Variable 2 (argument overlap), difference was along L1/L2 lines, but not proficiency level.

5.2 QUESTION 2: Do the Tendencies across the 6 Variables Using Different Language Backgrounds and Proficiency Levels Confirm that L2 Lexical Networks Are Generally Less Dense and L2 Writing Is Less Cohesive?

To answer Research Question 2, a simple and direct comparison of the results of this study were made with Crossley and McNamara's (2009) conclusions about L2 writing difference based on their own data from the 6 variables. Following Crossley and McNamara (2009), results were categorized from each variable as indicating with respect to L1 norms either L2 overuse, underuse or no significant difference (n.s.). Table 12 presents the results from both studies of the L2 data on each variable in terms of how they were shown to compare to native speakers.

Table 12. Comparison of results with earlier Coh-Metrix L2 research

	Crossley & McNamara (2009) L2 data	This study: L2 Low Proficiency	This Study: L2 High Proficiency
Variable 1 Causal content	Underuse	n.s.	Underuse
Variable 2 Argument overlap	Underuse	Overuse	Overuse
Variable 3 LSA sentence adjacent	Underuse	Overuse	Overuse
Variable 4 Noun hypernym	Underuse	n.s.	n.s.
Variable 5 Verb hypernym	Underuse	Underuse	Underuse
Variable 6 Freq of content words	Overuse	Overuse	Overuse

Table 12 shows that far from all of the tendencies found by Crossley and McNamara's (2009) on these 6 variables were generalizable features of L2 cohesion and lexical networks as they claimed. At least four of their tendencies, variable 1, 2, 3, and 4, did not correspond to the general tendencies of L2 writing found in this study

once different language backgrounds were included in the analysis. The tendency for overuse by L2 writers of frequently occurring content words was confirmed, similar to Mirzapour and Ahmadi (2011), as was the tendency for L2 writers to underuse abstract verb hypernyms. The lack of causal content as a cohesive device in L2 writing was supported in the L2 Higher Proficiency group, but was not true of the L2 Lower Proficiency learners. Regarding proficiency levels, as shown in Table 12, five out of the six variables indicate that both proficiency levels shared similar patterns of overuse, underuse or similar use of a feature compared to the native speaker norm.

6. Discussion

The results of this study reveal that five of the six Coh-Metrix measurements of cohesion and lexical network density (casual content, argument overlap, noun/verb hypernym, freq. content words) did not show a linear progression across proficiency levels toward the native speaker norms. One might conclude that either these variables do not seem to have the potential to track L2 development, or there is a lack of significant differences between proficiency groups as found previously by Chen (2008), Castro (2004) and Zhang (2002).

Although the Coh-Metrix did not detect significant proficiency level differences, it did consistently distinguish between L1/L2 writing. This supports research that argues cohesive devices mark L2 as non-native (Mirzapour & Ahmadi, 2011). This study indicates that L2 writers regardless of proficiency share similar lexical and cohesive patterns which mark it as distinct from L1 writing.

While validating that Crossley and McNamara's (2009) variables were able to analyze L1/L2 writing difference, four out of the six measures of L2 cohesion and lexical networks found in this study behaved differently than in their foundational study. This suggests that their conclusions about L2 cohesion and lexical network density were products of the single (Spanish) L1 background that formed their L2 data, and are not general patterns of L2 English. The results are reminiscent of Granger and Tyson (1996) who also found that the most common differences between NS writers and advanced EFL writers from a French language background were not generalizable features of advanced EFL writers across language backgrounds (see section 1.2).

7. Limitations and Future Research

The most important limitation of this study is that different data may have produced different conclusions as to whether the Coh-Metrix can measure L2 proficiency differences and lexical networks. Future research should include a wide range of written data at different proficiency levels, and representative of a variety of genres and dimensions of register (i.e. different tenors and fields). This will allow researchers to tease out any confounds that might be produced by a specific type of data and, when these are controlled, more definitively test the potential of the Coh-Metrix to track L2 development.

8. Conclusion

This study has shown that the computational tool Coh-Metrix can measure L1/L2 differences in cohesion and lexical network density, but not between high and low L2 proficiency levels. It showed that L2 writing contains more argument overlap, more semantic overlap, more frequent content words, less abstract verb hypernyms and less causal content than native speaker writing. The Coh-Metrix may prove an important and reliable new instrument for second language research. One can forse its use as pedagogical application that measures aspects of L2 academic writing, highlighting for teachers features in their student's writing that are markedly different from native speaker writing, which can then be addressed in the classroom.

References

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning, and Assessment*, *4*(3), 4-30.

Baayen, R., Piepenbrock, R., & Gulikers, N. (Eds.). (1995). *The CELEX lexical database* (CD-ROM). Philadelphia, Pennsylvania: Linguistic Data Consortium.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, *11,* 17-34. http://dx.doi.org/10.1017/S0272263100007816

Castro, D. (2004). Cohesion and the social construction of meaning in college students writing in L2 English. *Asia Pacific Education Review*, *5*(2), 215-225. http://dx.doi.org/10.1007/BF03024959

Chen, J. (2008). An investigation of EFL students' use of cohesive devices. *Asia Pacific Education Review*, *5*(2), 215-225.

Collins, P., & Hollo, C. (2010). *English grammar: an introduction*. New York: Palgrave-Macmillian.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33,* 497–505. http://dx.doi.org/10.1080/14640748108400805

Crossley, S., Greenfield, J., & McNamara, D. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, *42*(3), 475-493.

Crossley, S., & McNamara, D. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, *18*, 119-130. http://dx.doi.org/10.1016/j.jslw.2009.02.002

Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical proficiency using hypernymic relationships. *Language Learning*, *59*(2), 307-334. http://dx.doi.org/10.1111/j.1467-9922.2009.00508.x

Crossley, S., & McNamara, D. (2011). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Research in Reading*.

Cruse, D. (1986). *Lexical Semantics*. Cambridge: University of Cambridge Press.

Cumming, A. (2001). Learning to write in a second language: two decades of research. *International Journal of English Studies*, *1*(2), 1-23.

Ferris, D. (1994). Lexical and syntactic features in ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly, 28*(2), 414-420. http://dx.doi.org/10.2307/3587446

Ferris, D. (1994b). Rhetorical strategies in student persuasive writing: differences between native and non-native English speakers. *Research in the Teaching of English*, *28*(1), 45-65.

Graesser, A., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioural Research Methods, Instruments, and Computers*, *36*, 193–202. http://dx.doi.org/10.3758/BF03195564

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, *15*(1), 17-21. http://dx.doi.org/10.1111/j.1467-971X.1996.tb00089.x

Granger, S., Dagneaux, E., & Meunier, F. (2009). *The international corpus of learner English handbook and CD-ROM Version 2*. Louvain-la-Neuve, Belgium: Presses Universitaires de Louvain.

Granger, S., Sanders, C., & Connor, U. (2007). *LOCNESS: Louvain corpus of native English.* Belgium: Universitaires de Louvain.

Green, C., Christopher, E., & Mei, J. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: a corpus-based enquiry. *English for Specific Purposes*, *19*(2), 99-113. http://dx.doi.org/10.1016/S0889-4906(98)00014-3

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

Hudson, R. (2007). English dialect syntax in Word Grammar. *English Language and Linguistics, 11*(2), 383-405. http://dx.doi.org/10.1017/S1360674307002298

Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English. *RECL Journal of Language Teaching and Research*, *23*(2), 1-17. http://dx.doi.org/10.1177/003368829202300201

Kenkel, J., & Yates, R. (2009). The interlanguage grammar of information management in L1 and L2 developing writing. *Written Communication,* (26), 392-416. http://dx.doi.org/10.1177/0741088309341258

Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, *33*(4), 623-636. http://dx.doi.org/10.1016/j.system.2005.02.002

Lyons, J. (1995). *Linguistic Semantics: An Introduction*. Cambridge: University of Cambridge Press.

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2005, January 1). *Coh-Metrix version 1.4*. Retrieved from http//:cohmetrix.memphis.edu

McNamara, D., Ozuru, Y., Graesser, A., & Louwerse, M. (2006). Validating Coh-Metrix. In R. Sun, & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society*. Mahwah, NJ: Erlbaum.

Meara, P. (2005). Lexical frequency profiles: a Monte Carlo analysis. *Applied Linguistics*, *26*(1), 32-47. http://dx.doi.org/10.1093/applin/amh037

Meara, P. (2007). Simulating word associations in an L2: approaches to lexical organisation. *International Journal of English Studies*, *7*(2), 1-20.

Miller, G., Beckwith, A., Fellbaum, R., Gross, C., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3,* 235–244. http://dx.doi.org/10.1093/ijl/3.4.235

Mirzapour, F., & Ahmadi, M (2011). Study on Lexical Cohesion in English and Persian Research Articles. *English Language Teaching, 4*(4), 243–255. http://dx.doi.org/10.5539/elt.v4n4p245

Schiffrin, D. (1987). Why analyse discourse markers? *Discourse markers.* Cambridge: Cambridge University Press.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: the ESL research and its findings. *TESOL Quarterly*, *71*(4), 657-677. http://dx.doi.org/10.2307/3587400

White, L. (1987). Against comprehensible input: the input hypothesis and the development of second language competence. *Applied Linguistics*, *8*(2), 95-110.

Witte S., & Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication*, *32*(2), 189-204. http://dx.doi.org/10.2307/356693

Zhang (2004). Study on the Conjunction and Lexical Cohesion in College Students' English Compositions. *Journal of Kaifeng University*, *2*, 1-15.