



Lexical Complexity Development from Dynamic Systems Theory Perspective: Lexical Density, Diversity, and Sophistication

Reza Kalantari

Urmia University, Iran, r.kalantari@urmia.ac.ir

Javad Gholami

PhD, Corresponding Author, Urmia University, Iran, j.gholami@urmia.ac.ir

This longitudinal case study explored Iranian EFL learners' lexical complexity (LC) through the lenses of Dynamic Systems Theory (DST). Fifty independent essays written by five intermediate to advanced female EFL learners in a TOEFL iBT preparation course over six months constituted the corpus of this study. Three Coh-Metrix indices (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara & Graesser, 2012), three Lexical Complexity Analyzer indices (Lu, 2010, 2012; Lu & Ai, 2011), and four Vocabprofile indices (Cobb, 2000) were selected to measure different dimensions of LC. Results of repeated measures analysis of variance (RM ANOVA) indicated an improvement with regard to only lexical sophistication. Positive and significant relationships were found between time and mean values in Academic Word List and Beyond-2000 as indicators of lexical sophistication. The remaining seven indices of LC, falling short of significance, tended to flatten over the course of this writing program. Correlation analyses among LC indices indicated that lexical density enjoyed positive correlations with lexical sophistication. However, lexical diversity revealed no significant correlations with both lexical density and lexical sophistication. This study suggests that DST perspective specifies a viable foundation for analyzing lexical complexity.

Keywords: dynamic systems theory, lexical density, lexical diversity, lexical sophistication, lexical complexity development

INTRODUCTION

There has been a plenty of research on complex dynamic systems in physics, meteorology, and social sciences since the 1990s. The idea of complex systems was promoted by Larsen-Freeman (1997) in linguistic studies. Complex systems have attracted the attention of many scholars and researchers in language studies in the last few decades (e.g., Larsen-Freeman, 2011, 2014; Kyle, 2016; Verspoor, et al., 2011) and syntactic and lexical development research in particular (Bulte' & Housen, 2014; Caspi, 2010; Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Zheng, 2016).

Citation: Kalantari, R. & Gholami, J. (2017). Lexical Complexity Development from Dynamic Systems Theory Perspective: Lexical Density, Diversity, and Sophistication. *International Journal of Instruction*, 10(4), 1-18. <https://doi.org/10.12973/iji.2017.1041a>

Lexical complexity (LC) is presented in second language literature concerning lexical density, diversity/variability, and sophistication/rareness. It has been recognized as an indicator, diagnostic, and a major parameter for L2 learning, teaching, and research (Bulte' & Housen, 2014; Laufer, 1994; Wolfe-Quintero, Inagaki, & Kim, 1998). Many L2 academic writing studies have evaluated the scope to which these measures can be applied as reliable and valid determiners of learners' general language proficiency, particularly the quality of learners' writing performance, along with various criteria measures, including cohesion, coherence, organization, and discourse (Bulte' & Housen, 2014; Mazgutova & Kormos, 2015). The findings concerning LC development are inconclusive. Storch and Tapper (2009) operationalized lexical development as the percentage of words in Coxhead's (2000) academic word list (AWL) and reported a meaningful growth in lexical development after a short period of training, but Knoch, Rouhshad, and Storch (2014) and Deng, Lee, Varaprasad, and Lim (2010), who operationalized LC, similarly recorded no significant change in LC.

The study of lexical development in L2 writing has been an essential part of second language research in recent decades (e.g., Sasaki, 2007; Schmitt, 2010); however, little research has sought to establish links between lexical development and DST. Only in recent years, some scholars have addressed this connection (Bulté & Housen, 2014; Zheng, 2016). Drawing on language complexity and DST insights (Larsen-Freeman & Cameron, 2008; Bulté & Housen, 2014; Verspoor, de Bot, & Lowie, 2011), it is possible to introduce some perspectives into multi-constructed nature and complexity of L2 lexical proficiency. The present study closely expatiated on lexical characteristics of L2 learners' writing change during a longitudinal study, using nine different measures.

LITERATURE REVIEW

Dynamic Systems Theory

Study of second language learning has undergone numerous theoretical and practical changes in recent decades. One of the major and demanding aspects of second language research is language complexity. However, despite the interest engendered in a wealth of theoretical and empirical studies, there is no agreement on the definition of complexity and on how it has been characterized across or within studies, leading to terminological and conceptual confusion (Bulte' & Housen, 2014). In the last few decades, DST-inspired approaches to SLA have analyzed complexity in L2 writing via pursuing learners' written outputs over time to indicate the internal developmental dynamics of L2 complexity (e.g., Caspi, 2010; Kyle, 2016; Spoelman & Verspoor, 2010). The hallmark of this cross-disciplinary endeavour is complexity and nonlinearity of language development. Bulte' and Housen (2014) proposed complexity as a valid and basic L2 performance descriptor in L2 and L1 research as an indicator of language proficiency, development, and progress.

The dynamic trend of a complex system is confirmed further by availability of its resources. Considering the fact that resources are commonly constrained over various subsystems, some subsystems, as they support each other, interact as 'connected growers', while some others compete for limited resources as 'competitive growers',

connoting a trade-off relationship between the advanced grower and less advanced grower (van Geert, 1991). Based on DST, all systems are in constant change with chaotic variation, in which the systems only temporarily settle into “attractor states”. Systems constantly interact with their environment and reorganize themselves as a result of internal changes (de Bot and Larsen-Freeman, 2011). According to Larsen-Freeman and Cameron (2008), an attractor is a region in a system’s state space in which the system moves.

From DST perspective, language is a highly complex construct consisting of a set of interrelated variables/components, dimensions, and levels making it challenging to be independently evaluated. This means that language development is influenced by internal resources and external factors; consequently, changes in one system will have an impact over all other systems. Bulté and Housen (2012) proposed multidimensionality in L2 writing; they presented a taxonomic model of various components of language complexity as interpreted in L2 research, concluding that all different components of complexity may be evaluated across different language domains including the lexicon, syntax, and morphology (Bulté and Housen, 2014). They further pointed out that most L2 studies usually calculate only one or two complexity measures. Consequently, the multidimensional construct of complexity is reduced to one of its many possible operationalizations. Hence, complexity measurement acts are poor in content validity in extant L2 research. DST theory discusses that, since language is a complex dynamic system, the implementation of traditional methods to measure language development may not provide reliable or valid results. In order to predict how language development takes place, large amount of information is needed.

Following previous studies (Bulté & Housen, 2014; Lu, 2012; Read, 2000; Storch & Tapper, 2009; Zheng, 2016), LC is indicated as a multidimensional characteristic of language use including three interrelated components: lexical density, diversity, and sophistication. These measures are traditionally subsumed under a comprehensive construct of lexical richness. Bulté and Housen’s (2012) classified them as lexical diversity, while Jarvis (2013) suggested as consisting of volume, rarity, evenness, variability, dispersion, and disparity.

According to Johansson (2008), lexical density presents the proportion of lexical items in a text while lexical diversity measures different words used in a text. Lexical diversity or lexical variation is defined as the number of different words in a speech or writing sample with a determined length (Malvern, Chipere, Richards, and Duran, 2004). Lexical sophistication, also labelled as lexical rareness, is relatively advanced or rare proportion of words in learners’ writing (Read, 2000).

Previous studies have confirmed that LC has a multifaceted competence (Bulté & Housen, 2012, 2014; Schmitt, 2010; Zheng, 2016), and a good number of automated measures with varying reliabilities and validities have been developed to gauge LC indices (McNamara & Graesser, 2012; Lu, 2012; Cobb, 2000, to name but a few). To the best of our knowledge, most studies in this realm of inquiry have employed a single measure to account for LC, and their findings are based on one analysis instrument, both on data collection instrument and data analysis software. While building upon the

research tradition in these studies, investigating LC development of the same learners through multiple lenses for measuring this construct and the triangulation of their findings would probably yield more robust and accountable results. Moreover, it could put all these LC measurement instruments into test on what they measure and how they possibly differ from one another. Hence, in this study, multiple lexical measures were applied using the most up-to-date programs to address the multi-dimensional nature of lexical knowledge and included a select number of these indices in order to account for LC in the same learners' writing samples.

Another major incentive for this study was that, to the best of our knowledge, no study has been conducted on writing complexity development of Iranian EFL learners at a university level. A lot of Iranian post-graduate students attend TOEFL or IELTS preparation courses, one major module of which is concerned with their writing development. The participants in these courses are highly motivated and try their best to meet these two high stake proficiency exams' requirements. Given their immense investments in these courses from multiple perspectives, no systematic longitudinal investigation of their lexical development was found. Similarly, to our best knowledge, to what extent such courses could transform their lexical development has not been addressed so far. Therefore, the present inquiry set to investigate (a) how EFL learners develop their second language essay writing ability lexically over time, (b) whether there are any significant relationships between the sub-components (density, diversity, and sophistication) of LC.

METHOD

Design

The study employed longitudinal and descriptive-exploratory case study approach whereby the data were collected through the administration of a set of open-ended essay prompts. This design was felt most appropriate for this study because to explore language development assuming DST and changes over time, ELT researchers generally employ longitudinal design and case studies (e.g., de Bot, Lowie, Thorne & Verspoor, 2013; Caspi, 2010; Ortega & Byrnes, 2008; Salsbury, 2000; Verspoor, et al., 2011).

Participants

The participants of this study were five female learners of English who had a long experience of language learning in high school, private language schools, and university participate in the study. The sampling technique of this case study was purposeful (as was the case in Caspi, 2010; Salsbury, 2000; Verspoor, et al., 2011). The participants were intermediate-advanced English learners according to their TOEFL iBT total scores and its writing module. Their age range was between 24 and 37 (Table 1). They all had Azeri L1 background, studying non-English subjects at university. They were all postgraduate students, and their language proficiency level out of 120 in TOEFL iBT scale ranged between 70 and 90 (Mean score=79). Moreover, their scores in the writing module of the TOEFL iBT with one independent and one integrated writing task were found to vary from 17-22 out of 30 in TOEFL iBT scale.

They had studied English in the public education system and university for 10-12 years with limited hours of formal instruction. Meanwhile, these participants had attended private language schools for 3-12 years. In Iranian context, students study English at schools for two to four hours every week with locally developed textbooks which emphasize mostly grammar and reading comprehension with little attention to the development of writing and speaking skills. At the university, college students undertake general English and ESP courses with a heavy load of reading materials. Due to low efficiency of these formal language learning courses held in crowded classes with few resources, a considerable number of primary, secondary, and even tertiary students attend private language schools to improve their language proficiency communicatively and systematically (for review, see Naghdipour, 2016).

Table 1
Participants' profile

Name	Age	Major	Years of L2 learning experience		TOEFL iBT total score	TOEFL iBT writing score
			School and university	Private language school		
Shadi	27	Economics	10	3	71	17
Fatima	24	Medicine	10	12	89	22
Elham	37	Dentistry	10	10	87	21
Mahdis	25	Medicine	10	4	75	18
Yalda	27	Economics	10	4	73	18

The instructor was a highly-qualified male English teacher, with over 20 years of experience of teaching English at private and public schools. He completed his MA in TEFL from a reputable university in Iran. Thanks to his academic credentials and rich experience, he was assigned to run TOEFL iBT preparatory courses in the research site.

Writing tasks and data collection procedures

To explore the process of LC development, five intermediate to advanced learners of English who had enrolled in TOEFL iBT writing class were asked to take part in this study. The reason for selecting high proficiency learners of English was the fact that proficient learners have access to disparate resources, and self-organization can easily happen inside such a complex system (Verspoor et al., 2011). They attended the class twice a week in a period of six months and received instruction on both independent and integrated writing tasks in line with TOEFL iBT test.

The classroom instruction followed a step by step process-oriented and simulation-based L2 instruction through feeding, leading, showing, and throwing as main process options (McGrath, 1997). To begin with, the learners received instruction on key aspects of paragraph and essay writing such as topic sentence, thesis statement, paragraph unity, coherence, cohesion, logical progression of ideas, supporting one's ideas, and similar issues from the covered materials accompanied by teacher explanation, tips, and exemplification (feeding). Meanwhile, they were exposed to writing samples or templates with pre- or interactively-highlighted features of those model essays (showing). Later, they were engaged in some guided and staged writing practice

activities where they received on-the-spot assistance and scaffolding from their teacher, peers, and available resources such as their dictionaries (leading). Finally, as an integral component of the course, the learners were asked to compose a typical five-paragraph essay including introduction, main/body, and conclusion paragraph in the class on their own (throwing). To simulate the real TOEFL iBT exam conditions as was the primary aim of these learners in this course, the learners were required to rely on their own background knowledge and linguistic resources to craft these essays in typed format. On the topic prompt sheets, space was left for the learners to take notes if they wished. Using Dictionary was not allowed throughout this independent writing practice.

Every learner completed ten essays in word format over the study period. All essays with roughly two-week intervals were taken as sample corpus of the study. The instructor provided holistic and analytic written corrective feedback at his discretion on diverse aspects of the finished essays.

The main corpora of the study consisted of 50 essays. The participants were instructed to draft their essays observing the time limit and word length in TOEFL iBT test format. As they were getting prepared to sit official TOEFL iBT test, such simulated practice made sense for the participants and was honoured based on our field observations, anecdotal evidence, and their completed essays. The word length of the compositions ranged between 300-500 words (see Table 2). Consequently, the corpus consisted of a total of 18751 running words. The essays were chronologically ordered and saved in text file format. Because the writing task was computerized, the learners had access to their writing to correct mechanical and spelling errors so that the errors were very few in the writing essays. At the end, the topic prompts were deleted, proper nouns such as the names of geographical places or people's names in the essays were removed, and the main texts were imported into the LC analysers. The output results were importable to Excel and SPSS for further statistical analyses.

Table 2
Number and mean of words collected per participant

Name	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	MWL per essay	Total essay words
Shadi	273	254	333	324	169	316	325	300	303	274	287	2871
Fatima	541	504	383	363	364	433	409	392	334	370	409	4093
Elham	288	345	224	380	496	308	233	272	255	237	303	3341
Mahdis	361	370	276	253	375	267	382	345	638	359	362	3988
Yalda	336	534	375	352	382	409	531	387	311	436	405	4458
Mean	359	401	318	334	357	346	376	339	368	335	353	18751

Note: T: Time MWL: Mean Word Length

Data collection instruments

The learners were asked to answer 10 open-ended questions (essay topic prompts) and write 10 essays during six months with roughly two-week intervals. The genre of the

writing tasks was in academic writing register, all written under fairly similar circumstances. The essay topic prompts were all taken from the practice test books on TOEFL iBT similar to essay writing practices they experienced in the class.

To evaluate the learners' LC in terms of lexical density, diversity, and sophistication, nine lexical indices were selected to explore them through the participants' academic writing. Overall, three Lexical Complexity Analyzer indices (Lu, 2010, 2012; Lu & Ai, 2011), three Coh-metrix indices (Graesser, et al., 2004; McNamara & Graesser, 2012), and four Vocabprofile indices (Cobb, 2000) were selected to account for LC. To put it differently, two indices were selected to define lexical density, three indices to define lexical diversity, and four indices to define lexical sophistication. Table 3 displays the nine indices representing LC.

Lexical density was measured through two analyzers: lexical complexity analyzer (LCA) (Lu, 2010, 2012; Lu & Ai, 2011) and Vocabprofile software (Cobb, 2000).

Four lexical measures served as the indices of lexical diversity, namely the measure of textual lexical diversity (MTLD), vocabulary diversity (Vocd-D), Uber Index (Uber), and squared verb variation (SVV).

Three indices of academic word list (AWL) and Beyond-2000 scores (B-2000) and log frequency (LogF) of content words were used to gauge lexical sophistication. All instruments were valid and reliable software. For instance, Coh-Metrix can reach a reliability of 0.92 in texts with a particular genre) McNamara & Graesser, 2008). Vocabprofile is also a reliable measure of lexical complexity with reliability of more than 0.75 for its different indices (Abbasian & Shiri Parizad, 2011). Lexical Complexity Analyzer also correlates strongly with the raters' judgments of the quality of ESL learners' oral narratives from moderate to high ($r=0.53$ to 0.76) for lexical density, diversity, and sophistication (Lu, 2012).

Table 3
Software specifications to measure lexical complexity

Lexical Complexity	Indices	Definition	Softwares
Density	Lexical Density (LD-LCA)	Content word ratio	Lexical Complexity Analyzer
	Lexical Density (LD-VP)	Content word ratio	Vocabprofile software
Diversity	Uber Index (Uber)	The proportion of the squared number of log to the whole number of log in the text.	Lexical Complexity Analyzer
	Squared Verb Variation (SVV)	The proportion of the squared number of verb types to the whole number of verbs in the text.	Lexical Complexity Analyzer
	Measure of Textual Lexical Diversity (MTLD)	The average length of sequential word strings in a text which maintain a given TTR value.	Coh-Metrix 3
	Vocabulary Diversity (Vocd-D)	A mathematical transformation of the standard type-token ratio (TTR) which	Coh-Metrix 3

			reduces the intervening impacts of text length and indicates the degree of words' repetition in a text.	
Sophistication	Academic Word Length (AWL)	Word	A list of 570 frequent words in an academic context.	Vocabprofile software
	Beyond-2000(B-2000)		The Beyond-2000 values calculated by subtracting K1 and K2 ratios from 100%.	Vocabprofile software
	Content Word Frequency (LogF)	Log	The average of the log frequency of content words in the text	Coh-Metrix 3

Data analysis

After analysing the texts via lexical analysers, the researchers subjected the results to descriptive statistics. In order to determine whether lexical development occurred over time with regard to the lexical indices of interest, repeated measure analysis of variance (RM ANOVA) statistics was performed. Meanwhile, to find the relationships among lexical indices, Pearson product-moment correlation test was employed.

FINDINGS

To measure learners' lexical development, (RM ANOVA) statistics were conducted using LC indices (LD-LCA, LD-VP, Uber, SVV, MTL, Vocd-D, AWL, B-2000, LogF). Tables 4 to 6 provide descriptive statistics on lexical indices. Before conducting RM ANOVA, normality of the data was examined.

Table 4

Means and standard deviations of lexical density indices

IND/T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
LD- LCA	.51	.49	.51	.51	.52	.54	.051	.51	.54	.53
SD	.026	.041	.034	.030	.048	.025	.032	.043	.036	.041
LD- VP	.50	.48	.51	.51	.51	.53	.50	.53	.55	.52
SD	.021	.041	.052	.032	.052	.033	.034	.034	.034	.052

The descriptive statistics of lexical density indices in Table 4 reveals that lexical density means measured by Lexical Complexity Analyzer (LD-LCA) changed slightly from .51 to .53. Similarly, it changed from .50 to .52 measured by Vocabprofile software (LD-VP). However, it fluctuated between .49 and .54 by LD-LCA analyzer and between .48 and .55 by LD-VP analyzer within time intervals.

Table 5

Means and standard deviations of lexical diversity indices

IND/T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
UBER	19.5	20.3	20.0	20.5	19.9	20.9	20.2	20.6	20.6	20.0
SD	2.19	2.34	1.02	1.99	2.85	2.83	1.91	2.26	3.64	2.07
SVV	21.9	23.5	24.5	26.1	24.9	23.5	24.5	23.7	22.5	24.6
SD	6.62	7.18	6.19	1.77	10.1	4.28	5.50	2.39	9.96	6.78
MTLD	77.4	96.2	69.	83.7	88.1	98.7	95.2	93.6	97.7	91.7
SD	21.1	17.7	10.	16.7	25.6	26.6	10.9	12.2	19.3	22.2
Vocd-D	79.8	99.0	85.	86.8	89.8	93.4	91.5	92.2	92.2	92.8
SD	13.8	18.5	4.21	15.9	21.1	22.6	9.56	15.0	22.9	6.80

Table 5 provides descriptive statistics on lexical diversity indices measured by lexical diversity instruments, namely the measure of textual lexical diversity (MTLD), vocabulary diversity (Vocd-D), Uber Index (Uber), and squared verb variation (SVV). As seen in Table 5, Uber index went up from 19.5 to 20, SVV rose from 21.9 to 24.6, MTLD witnessed an increase from 77.4 to 91.7, and finally Vocd-D increased from 79.8 to 92.8 over six months.

Table 6

Means and standard deviations of lexical sophistication indices

IND/T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
AWL	4.58	3.68	3.54	3.18	5.14	7.51	7.57	6.32	8.27	8.15
SD	1.85	1.89	1.22	1.18	1.54	3.47	2.21	3.39	1.41	3.05
B-2000	6.85	7.85	8.61	9.08	10.6	9.58	13.4	12.8	9.83	17.8
SD	2.29	3.22	2.00	1.67	2.98	1.87	4.22	3.66	4.99	3.30
LogF	2.93	2.91	2.85	2.85	2.97	2.90	2.93	2.92	2.98	3.05
SD	.21	.40	.34	.32	.14	.29	.27	.36	.05	.08

Finally, Table 6 presents the descriptive statistics on lexical sophistication measures. As the table shows, Academic Word Length (AWL) changed from 4.58 to 8.15; Beyond-2000 (B-2000) increased from 6.85 to 17.8, and Content Word Log Frequency (LogF) saw an improvement from 2.93 to 3.0 over this period.

The next issue was to consider Sphericity Mauchly's Test of Sphericity. As nothing is known about Sphericity at all, the hypothesis that the assumption of Sphericity has not been violated is accepted, hence, the Greenhouse-Geisser correction was used. Table 5 summarizes the results of effects of RM ANOVA, using Greenhouse-Geisser measure. The results yielded a positive significant difference between time and mean values of only AWL and B-2000 (indicators of lexical sophistication): (AWL, $P < .001$, $\eta^2 p = .389$; B-2000, $P < .001$, $\eta^2 p = .347$). No significant difference was observed between time and means of the remaining seven indices (LD-LCA, LD-VP, Uber, SVV, MTLD, Vocd-D, and LogF).

Table 7

Repeated measure ANOVA results for lexical indices at time intervals

Indices	Type III Sum of Squares	Mean Square	F	Sig.	Partial Eta Squared
LD- LCA	.010	.005	1.399	.300	.235
LD-VP	.016	.006	1.748	.215	.309
UBER	7.911	2.886	.196	.882	.305
SVV	68.693	24.641	.170	.904	.310
MTLD	4226.406	1575.650	1.772	.214	.298
Vocd-D	1220.529	433.464	.556	.645	.313
AWL	181.867	51.941	3.979	.026	.389
B-2000	471.535	150.847	4.971	.016	.347
LogF	.169	.113	.490	.584	.167

As the results of repeated measures analysis indicate, a significant difference was only detected in lexical sophistication development. There were significant developmental differences between time spent in writing English and AWL and B-2000. Nevertheless, lexical density and diversity tended to flatten during a six-month writing instruction course.

To understand the relationships among LC, the indices were correlated separately using Pearson correlation.

Table 8
Means and standard deviations for lexical complexity indices

Lexical Indices	Std. Deviation	Mean
LD- LCA	.5202	.025
LD-VP	.5126	.020
UBER	22.34	1.623
SVV5	24.00	1.829
MTLD	89.19	11.247
VOCD-D	90.35	6.792
AWL	6.17	.713
B-2000	12.07	2.89
LogF	2.93	.199

Table 8 illustrates descriptive analysis of the relationships among nine lexical complexity indices including means and standard deviations.

Table 9
Correlations among lexical complexity indices

	LD-LCA	LD-VP	Uber	SVV	MTLD	Vocd-D	AWL	B-2000	LogF
LD- LCA	1	.897*	.412	-.349	.680	.340	.914*	-.352	-.728
LD- VP		1	.039	.491	.565	.207	.576	.030	.073
Uber			1	.058	-.491	.468	.158	.797	-.386
SVV				1	.446	.667	.888*	.149	.226
MTLD					1	.663	.451	.097	-.502
Vocd-D						1	-.069	.740	-.421
AWL							1	-.688	-.787
B-2000								1	.200
LogF									1

*. Correlation is significant at 0.05 (2-tailed).

Table 9 shows the analysis of the relationships among nine lexical indices and Pearson correlations. As these figures show, there are correlations among only a few lexical indices, although the correlation coefficients among them are remarkably high ($r=.89$, $.91$, $.88$). Among these, LD-LCA and LD-VP as two indicators of lexical density were significantly correlated ($r=.897$). Moreover, LD-LCA was highly correlated with AWL ($r=.91$). Uber was strongly correlated with Vocd-D ($r=.88$). There were no significant correlations among other LC indices.

DISCUSSION

This study primarily explored Iranian EFL learners' LC development from DST perspective and then investigated possible correlations among lexical indices. An improvement was only found in two indices of lexical sophistication. However, such an improvement was not observed in other indices which plateaued out over time. Meanwhile, correlation analyses of lexical indices revealed positive relationships between two indicators of lexical density, namely LD-LCA and LD-VP, indicating that the employed lexical density analysers were highly reliable and valid (Table 4). LD-LCA enjoyed a positive correlation with AWL as an indicator of lexical sophistication and Vocd-D indicated positive correlations with Uber.

Lexical sophistication development and LC indices relationships manifested in the current analysis by a significant increase corroborate the findings of Zheng (2016) who demonstrated divergent developmental patterns in different aspects of LC. However, unlike Zheng's finding which revealed increases in lexical sophistication and diversity, progress was found only in the case of lexical sophistication. The increased lexical sophistication mirrors the findings of previous studies (Jarvis, 2002; Malvern, et al., 2004; Storch & Tapper, 2009; Zheng, 2016), excluding the fact that they also found a growth in lexical diversity. The present findings are also consistent with those of Linnarud (1986) who noted significant differences in lexical sophistication regarding writings by native English speakers and Swedish learners of English and also Laufer (1994) who found significant differences in lexical sophistication between pre-and post-writings performed by two university classes' students. However, this change is inconsistent with Bulte' and Housen (2014), Knoch et al. (2015), and Storch and Tapper (2009). For instance, in Bulte' and Housen's (2014) study, only one out of the three LC measures, namely lexical sophistication, underwent an increase from Time 1 to Time 3, but that increase was not meaningful, and the scores decreased slightly and non-significantly over time with respect to lexical density and lexical richness. Riazi (2016) revealed significant differences in MTLTD and LogF related to lexical sophistication according to the task type and task similarity. Polat and Kim (2014) reported that the development occurred only in the lexical diversity of the participants.

One possible explanation for these contradictory results and many others in the literature is that unlike supportive and parallel development of syntactic complexity among high proficiency learners (authors, n.d.), LC development, at least for non-native users, is less likely to develop in parallel in terms of diversity, density, or sophistication over a period of six months as was the case in the current study. It appears that those lexical indices are separate entities with their own distinct developmental paths as the findings in this study demonstrate, supporting the idea that they are different aspects of L2 proficiency repertoire.

The following interpretations could be made to account for non-parallel development of different aspects of LC as was the case in this study. One plausible explanation is that the learners in the study attended these courses with the hope of boosting their scores in TOEFL iBT course. This incentive and the backwash effect of the exam could have influenced the course of their lexical development. We are of the conviction that among

various indices of vocabulary, lexical sophistication is surgically one which presumably lends itself to be upgraded more easily than the others in a short period of time, and thus has received more attention. It is probable that these students sought to learn a limited set of sophisticated words, as they are found in GRE word lists. These learners might have tried to squeeze such words within their essays, thereby boosting their lexical sophistication.

In addition, the exam-oriented context of the course may partially explain the findings. Unlike other general English courses in which all aspects of writing complexity are emphasized both in instruction and course materials, in high stake exam preparation courses, at least in Iranian context as far as we know, most learners' attention is drawn to the development of a good orientation to the exam itself, its layout, task type, and simulated practice. As the exam itself is on the spotlight in these courses, there are very few opportunities for the course participants to enrich their competencies in all indices of vocabulary including density and diversity in tandem with sophistication.

Another factor that could contribute to the obtained results of this study is related to the type of feedback the learners have received from their teachers. As we reviewed the written corrective feedback these learners received from their course instructor, we noticed that the bulk of the feedback was concerned with highlighting their syntactic errors, whereas the number of lexically oriented feedback instances was very low. Our own written corrective feedback practice also corroborates this observation. Although we did not quantify the proportions of syntactic and lexical development feedback types and are not in a position to offer conclusive results on this issue, we think that this discriminatory focus between syntax and lexicon may play a role in why lexical development tended to flatten over the course of this writing program in most of its indices.

Another plausible justification for asynchronous LC developmental patterns refers to the co-adaptive interactions between dynamic systems' sub-systems which possibly bring about a pattern of equilibrium; the occurrence of lexical plateau shows the arrival of the attractor state, which is in line with Zheng (2016) and de Bot and Larsen-Freeman (2011) who the view that processing system is limited by constrained recourses, making it difficult for learners to attend simultaneously to disparate aspects of complexity. If the results are considered from this perspective, the lexical plateau may reflect an attractor state due to co-adaptation, rather than a compulsory failure or a final phase in the process of development. In this study, lexical sophistication developed, while lexical density and diversity stabilized.

CONCLUSION

EFL learners' LC was assessed through nine indices, and correlations among its sub-components were drawn. Lexical density was measured through two analysers: lexical complexity analyser and Vocabprofile. Lexical diversity was quantified by the Uber's index, squared verb variation, measures of textual lexical diversity, and Vocd-D, and lexical sophistication was estimated by the academic word list, Beyond-2000 scores, and Log frequency of content words. The results indicated development only in lexical

sophistication, while other aspects of LC reached plateau. The major advantage of this research was its longitudinal investigation of LC through employing three distinct analysers, making the findings more accountable compared with those obtained through a single measure.

The present study offers implications for L2 teaching and learning. The development of only two LC indices out of nine underscores the fact that various aspects of one's lexical competence do not develop uniformly. Language teachers are recommended to note that L2 lexical development may remain flat in some dimensions, while others could develop well and at a faster pace or in a shorter time (Zheng, 2016). Moreover, the three dimensions of LC with nine indicators were not strongly correlated, suggesting indeed different constructs. None of the lexical diversity indices were correlated with the indices related to lexical sophistication.

Another noteworthy methodological finding is that employing multiple analysers and indices for lexical development in an extended period of time is fruitful and promising. The analysis carried out in the DST framework shows the complexity, interconnectivity, and/or independency of the L2 lexical systems. The accumulation of the findings through employing multiple measures demonstrated the significance of neglecting one-size-fits-all measure of L2 complexity as voiced by Bulté and Housen, (2014). Employing triangulated and automated measures to analyse LC can assist researchers and teachers to evaluate LC far more quickly, automatically, and reliably. In validating testing procedures and their interpretation, McNamara (2006) reminds us to have plenty of evidence in order to judge reliably and validly, unless the interpretation is unsound and faulty. Using multiple analysers, this study provided a better understanding of LC construct, rendering us the opportunity to create reliable tools for lexical instruction and assessment. Characterizing how lexicons develop over a long-term period can assist material and curriculum designers to provide lexical repertoires that match learners' abilities. These automated measures can also be used as a diagnostic tool to identify students' lexical development in order to improve their writing skill.

As with most studies, the present study has its own limitations. In spite of the considerable indices applied and the use of reliable softwares, these analysers may still be too tough to precisely and completely specify L2 learners' lexical production and may not fully capture one's lexical development. The same holds true regarding subcomponents of LC. LC presented through three hidden variables of lexical density, diversity, and sophistication was taken into account. The insights obtained from the study raise some interesting questions on how other writing elements, i.e. cohesion and coherence, interact with lexical complexity in L2 writing development.

REFERENCES

- Abbasian, G. R., & Shiri Parizad, M. (2011). Validation of lexical frequency profiles as a measure of lexical richness in written discourse. *Journal of Technology & Education*, 5(3), 1-7.
- Bulte', B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Ed), *Dimensions of L2 performance and proficiency*

- *Investigating complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26(4), 42–65.
- Caspi, T. (2010). *A Dynamic perspective on second language development* (Unpublished doctoral dissertation). University of Groningen, Netherlands.
- Cobb, T. (2000). Web Vocabprofile, accessed October, 2016 from <http://www.lextutor.ca/vp>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Davis, C. J. (2005). N-Watch: A program for deriving neighbourhood size and other psycholinguistic statistics. *Behaviour Research Methods*, 37(1), 65–70.
- de Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In M. H. Verspoor, K. de Bot & W. Lowie (Ed), *A dynamic approach to second language development: Methods and techniques* (pp. 5–24). Amsterdam: Benjamins.
- de Bot, K., Lowie, W., Thorne, S.L., & Verspoor, M. (2013). Dynamic systems theory as a comprehensive theory of second language development. In M. Mayo, M. Gutierrez-Mangado and M. Adrián (Ed), *Contemporary Approaches to Second Language Acquisition* (pp. 199–220). Amsterdam: John Benjamins.
- Deng, X., Lee, K. C., Varaprasad, C., & Lim, M. L. (2010). Academic writing development to ESL/EFL graduate students in NUS. *Reflections on English Language Teaching*, 9(2), 119–138.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioural Research Methods, Instruments, and Computers*, 36(2), 193–202.
- Jarvis, S. (2002). Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing*, 19(1), 57–84.
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63, 87–106.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Lund University, Dept. of Linguistics and Phonetics Working Papers*, 53, 61–79.
- Knoch, U., Rouhshad, A., & Storch, N. (2014). Does the writing of undergraduate ESL students develop after one year of study in an English-medium university? *Assessing Writing*, 21(1), 1–17.
- Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, 28, 39–52.

- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Unpublished doctoral dissertation). Georgia State University.
- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141–165.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese Learners of English. *Applied Linguistics*, 27(4), 590–919.
- Larsen-Freeman, D. (2011). A complexity theory approach to second language development/acquisition. In D. Atkinson (Ed). *Alternative Approaches to Second Language Acquisition* (pp. 48–72). New York: Routledge.
- Larsen-Freeman, D. (2014). Saying what we mean: Making the case for second language acquisition to become second language development. *Language Teaching*, 48(4), 491-505.
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistic*. Oxford University Press.
- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, 25(2), 21–33.
- Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Lund, Sweden: CWK Gleerup.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496.
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190-208.
- Lu, X., & Ai, H. (2011). Synlex: Lexical complexity analyzer complexity analyzer. L2 syntactical complexity analyzer. (Computer software). Retrieved from <http://aihaiyang.com/software/>
- Ortega, L., & Byrnes, H. (2008). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (Ed), *The longitudinal study of advanced L2 capacities* (pp. 3-20). New York: Routledge.
- Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in intensive English for academic purposes program. *Journal of Second Language Writing*, 29, 3-15.
- McCarthy, P., & Jarvis, S. (2010). MTL, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392.
- McGrath, I. (1997). Feeding, Leading, Showing, Throwing: "Process Choices in Teacher Training and Trainer Training", In I. McGrath (Ed) *Learning to Train: Perspectives on the development of language teacher trainers*. Prentice Hall: Hemel 13 Hempstead.

- McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language assessment quarterly*, 3(1), 31-51.
- McNamara, D. S., & Graesser, A. C. (2008). In P. M. McCarthy & C. Boonthum (Ed), *Applied natural language processing: Identification, investigation, and resolution*. Hershey, PA: IGI Global.
- McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy, & C. Boonthum-Denecke (Ed), *Applied natural language processing: Identification, investigation, and resolution* (pp. 188–205). Hershey, PA: IGI Global.
- Naghdipour, B. (2016). English writing instruction in Iran: Implications for second language writing curriculum and pedagogy. *Journal of Second Language Writing*, 32, 81-87.
- Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: a longitudinal case study of advanced untutored development. *Applied Linguistics*, 35(2), 184–207.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Riazi, A. M. (2016). Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features. *Assessing Writing*, 28, 15–27.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Salsbury, T. (2000). *The acquisitional grammaticalization of unreal conditionals and modality in L2 English: A longitudinal perspective* (Unpublished doctoral dissertation). Indiana University, Bloomington.
- Sasaki, M. (2007). Effects of study-abroad experience on EFL writers: A multiple-data analysis. *The Modern Language Journal*, 91 (4), 602-620.
- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in the development of accuracy and complexity: A longitudinal case study on the acquisition of Finnish. *Applied Linguistics*, 31(4), 532–553.
- Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes*, 8(3), 207-223.
- Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, 98(1), 3-53.
- Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214–231.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Hawaii: University of Hawaii.
- Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, 56, 40-53.

Turkish Abstract**Dinamik Sistem Teorisi Perspektifinden Sözcüksel Karmaşıklık Geliştirme: Sözcüksel Yoğunluk, Çeşitlilik ve Sofistike**

Bu boylamsal durum çalışmasında, İran EFL öğrencilerinin, objektif bir yolla Dinamik Sistem Teorisinin (DST) sözcüksel karmaşıklığını (LC) araştırılmıştır; TOEFL iBT hazırlık dersinde altı aydan uzun bir sürede beş orta ve ileri seviyedeki bayan EFL öğrencileri tarafından yazılan elli bağımsız makale, bu çalışmanın temelini oluşturmuştur. Akademik Sözcük Listesi ve Beyond-2000'de sözcük karmaşıklığının göstergesi olarak zaman ve ortalamalar arasında pozitif ve anlamlı ilişkiler bulunmuştur. Geriye kalan altı yedi LC indeksinin, bu yazma programının gidişatını aşma eğiliminde olduğu görülmüştür. Bu çalışma, DST bakış açısının, sözcüksel karmaşıklığı analiz etmek için uygun bir temel oluşturduğunu önermektedir.

Anahtar Kelimeler: dinamik sistem teorisi, sözcük yoğunluğu, sözcük çeşitliliği, sözcük karmaşıklığı

French Abstract**Développement de Complexité Lexical de Perspective de Théorie de Systèmes Dynamique : Densité Lexicale, Diversité et Sophistication**

Cette étude de cas longitudinale a exploré la complexité lexicale des apprenants d'EFL iraniens (LC) par les lentilles de Théorie de Systèmes Dynamique (DST). Cinquante essais indépendants écrits par cinq intermédiaire à la femme avancée des apprenants d'EFL dans un TOEFL iBT le cours de préparation plus de six mois ont constitué le corpus de cette étude. Les relations positives et significatives ont été trouvées entre le temps et des valeurs moyennes dans la Liste de Mot Universitaire et Au-delà - 2000 comme les indicateurs de sophistication lexicale. Le maintien sept indices de LC, la non perte à la signification, a eu tendance à aplanir au cours de ce programme d'écriture. Cette étude suggère que la perspective DST spécifie une fondation viable pour analyser la complexité lexicale.

Mots Clés: théorie de systèmes dynamique, densité lexicale, diversité lexicale, sophistication lexicale

Arabic Abstract**تنمية التعقيد اللغوي من منظور النظريات الديناميكية: الكثافة المعنوية، التنوع، والتطور**

استكشفت دراسة الحالة الطولية هذه التعقيد المعجمي الإيراني لـ EFL من خلال عدسات نظرية الأنظمة الديناميكية (DST). وقد شكلت خمسين مقالة مستقلة كتبها خمسة متعلمين متوسطين إلى متقدمين في اللغة الإنجليزية كلغة أجنبية في دورة إعداد توفل iBT على مدى ستة أشهر مجموعة من هذه الدراسة. ووجدت علاقات إيجابية وهامة بين الوقت والقيم المتوسطة في قائمة الكلمات الأكاديمية وما بعدها 2000 كمؤشرات للتطور المعجمي. المؤشرات السبعة المتبقية من LC ، التي نقل أهمية، تميل إلى تسطيح على مدى هذا البرنامج الكتابية. وتشير هذه الدراسة إلى أن منظور DST يحدد أساساً قابلاً للتطبيق لتحليل التعقيد المعجمي.

الكلمات الرئيسية: نظرية النظم الديناميكية، الكثافة المعجمية، التنوع المعجمي، التطور المعجمي

German Abstract**Lexikalische Komplexität Entwicklung aus dynamischen Systemen Theorie Perspektive: Lexikalische Dichte, Vielfalt und Raffinesse**

Diese Längsfallstudie untersuchte die lyrische Komplexität (LC) der iranischen EAZ-Lernenden durch die Linsen der dynamischen Systemtheorie (DST). Fünfzig unabhängige Aufsätze, die von fünf Fortgeschrittenen zu fortgeschrittenen weiblichen EAZ-Lernenden in einem TOEFL iBT Vorbereitungskurs über sechs Monate geschrieben wurden, stellten das Korpus dieser Studie dar.

Positive und signifikante Beziehungen wurden zwischen Zeit- und Mittelwerten in der akademischen Wortliste und jenseits von 2000 als Indikatoren der lexikalischen Raffinesse gefunden. Die verbleibenden sieben Indizes von LC, die der Bedeutung unterliegen, tendierten dazu, im Laufe dieses Schreibprogramms zu glätten. Diese Studie deutet darauf hin, dass die DST-Perspektive eine tragfähige Grundlage für die Analyse der lexikalischen Komplexität festlegt.

Schlüsselwörter: dynamische systemtheorie, lexikalische dichte, lexikalische vielfalt, lexikalische raffinesse

Malaysian Abstract

Pengembangan Kompleks Lexikal dari Perspektif Teori Sistem Dinamik: Ketumpatan Lexikal, Kepelbagaian, dan Kecanggihan

Kajian kes longitudinal ini meneroka kompleksiti leksikal (LC) pelajar EFL Iran melalui perspektif Teori Sistem Dinamik (DST). Lima puluh esei bebas yang ditulis oleh lima pengantara untuk pelajar EFL wanita dalam kursus penyediaan TOEFL iBT selama enam bulan merupakan korpus kajian ini. Hubungan positif dan signifikan ditemui antara nilai masa dan min dalam Senarai Akademik dan Beyond-2000 sebagai petunjuk kecanggihan leksikal. Baki tujuh indeks LC, yang kurang penting, cenderung meratakan sepanjang program penulisan ini. Kajian ini menunjukkan bahawa perspektif DST menentukan asas yang berdaya maju untuk menganalisis kerumitan leksikal.

Kata Kunci: teori sistem dinamik, kepadatan leksikal, kepelbagaian leksikal, kecanggihan leksikal

Russian Abstract

Развитие Лексической Сложности из Теории Динамических Систем Перспектива: Лексическая Плотность, Разнообразие и Изощренность

В этом исследовании изучалась лексическая сложность английского языка в качестве иностранного языка для иранских студентов, используя основы теории динамических систем. Пятьдесят независимых статей, написанных пятью женщинами-студентками EFL на протяжении шестимесячного курса подготовки к TOEFL iBT представляют собой основу данного исследования. Позитивные и значимые отношения были найдены между временными и средними значениями в Academic Word List и Beyond-2000. Оставшиеся семь индексов LC, не достигая значимости, как правило, сглаживались в ходе этой программы написания. Это исследование предполагает, что перспектива DST указывает на достаточную основу для анализа лексической сложности.

Ключевые Слова: теория динамических систем, лексическая плотность, лексическое разнообразие, лексическая изощренность