The Development of Intra-Individual Variability in Academic Writing: A Study on Lexical Diversity and Lexical Sophistication

Muzakki Bashori*

Vocational High School of Wisudha Karya, Kudus, 59319, INDONESIA

Abstract
This study is aimed at finding out (1) whether Lexical Diversity (LD) and Lexical Sophistication (LS) can provide useful insights into the development of academic writing by tracing the interaction between the intra-individual variability in relation to a Dynamic Systems perspective, and (2) whether the supportive interaction between LD and LS can be recognized from the writing development. Twelve academic writing samples written over a 5-year period (2010-2015) by an Indonesian learner of English were employed as longitudinal data. Several tools designed by van Geert and van Dijk (2002), Peltier (2009), and Steinkrauss (2016) were used to analyze the dynamic patterns of language development. The results showed that the development of intra-individual variability in academic writing is in line with the Dynamic Systems Theory as it indicates that the developmental process between the two growers is complex, non-linear, self-organized, unpredictable, revealing attractor states, and constantly changing. The supportive growth movement emerges as the result of the interaction between variables. Finally, it can be concluded that variability is a source of development. Learners might need to be aware of their unique learning trajectory in order to maintain a more stable linguistic development.

Keywords: Academic writing, dynamic systems perspective, lexical diversity, lexical sophistication, variability.

* Corresponding author, email: muzakkibashori90@gmail.com


Received June 1, 2020; Revised March 08, 2021; Accepted March 15, 2021; Published Online May 3, 2021

https://doi.org/10.24815/siele.v8i2.16843
1. INTRODUCTION

The Dynamic System Theory (DST) posits that the presence of variability is a source of development, from which one’s language processing is seen as a non-linear, complex, individual, and dynamic trajectory (Lowie, 2013). Showing the peaks and troughs within the process is one of the normal characteristics of the system (de Bot et al., 2005). Those moments of progress and regress are surely unavoidable in accordance with the assumption of the dynamic interaction of subsystems in the developmental data. Several studies have also sought to discover if the variability being examined relates to each other meaningfully over time and can be gauged with some special measurements (Spoelman & Verspoor, 2010; van Geert & van Dijk, 2002; Verspoor et al., 2008; Verspoor et al., 2012). Therefore, longitudinal research is considered to best fit in explaining the language development of a learner.

This study attempts to investigate intra-individual variability, specifically in terms of (1) Lexical Diversity (LD) and (2) Lexical Sophistication (LS) in academic writings of an Indonesian learner of English over five years (2010-2015). The methods and tools developed by van Geert and van Dijk (2002), Peltier (2009), and Steinkrauss (2016) were employed to indicate the dynamic development of the two growers. This research aims to give worthwhile insights whether (1) LD and LS can provide useful insights into the development of academic writing by investigating the interaction between the intra-individual variability in relation to a Dynamic Systems perspective, and (2) whether the supportive interaction between LD and LS can be recognized from the writing development.

This study is divided into five sections. The first section aims at introducing the topic of discussion. The second section provides some literature reviews on a Dynamic Systems perspective as the background of study and also includes previous research related to the key topic. The third section deals with the research methodology used for this study. The fourth section discusses the results of observation towards two aspects of variability and their interaction, and provides discussion of the findings. Finally, the last section summarizes the research findings as well as presents briefly the possibility to conduct further research within the scope of this study.

2. LITERATURE REVIEW

2.1 (Second) Language Development as a Dynamic System

Where there is variability, there is development. (Lowie, 2013, p. 21)

The Dynamic Systems Theory (DST) claimed its position to be inherently compatible to understand better the notion of development. It has been discussed that this theory deals with the assumption that the development of a system is iterative, difficult to predict, individual, interacting, self-organized, unstable, fluctuative, and highly variable (Lowie, 2013). Larsen-Freeman (1997) pioneered the connection between this theory and second language development through her publications, which triggered a multitude of research on that approach since that time continuously. Moreover, research on second language development is assumed to be even more complex and interesting than the first language due to many factors or subsystems to
take into account, for instances, prior knowledge of another language, aptitude, feedback on language, learning motivation, and type of instruction (O’Grady, 2008), and exposure to the language (de Bot et al., 2005). Thus, viewing the process of the second language (L2) development through observing L2 learner’s academic writings from a Dynamic Systems perspective will considerably provide supporting evidence proving that the learner’s developmental trajectory is surely dynamic.

The Dynamic Systems Theory (DST) is also in line with Emergentism, for its principles complement the construct that a structure that is emergent, is taken to be constantly open and in flux (Hopper, 1998, in Spoelman & Verspoor, 2010). Both approaches believe that the resources for a system to grow in such a development should be available as a prerequisite and are limited. The term ‘resource’ refers to a complex of internal or external factors possibly affecting or being used by a learner (van Geert, 1994). In other words, it can be inferred that these views are a matter of change in individuals, which implies that interaction among subsystems over time is of considerable importance, and all internal and external factors are absolutely interconnected and dependent on one another. Consequently, the stages of development can be identified by scrutinizing continuous and discontinuous variability patterns in each individual grower. The variability patterns between growers can be supportive, competitive, or pre-required (Verspoor et al., 2011). However, it is important to note that the patterns may vary as this developmental process is dynamic, non-linear, and even unpredictable, depending on what types of growers they are and how they interact over time.

This study, therefore, carefully examines intra-individual variability in longitudinal data to obtain a clear understanding of second language development. Previous studies have been conducted in various and creative ways to shed a light on this topic of discussion as the following part explains. This part also serves as one of the underlying points why the author chose particular variability to detect the presence of development.

### 2.2 Previous Studies

Verspoor et al. (2017) argued that the second language development of a learner can be traced by analyzing his/her writings in the target language within a certain period of time. The way their research measured variability eventually enabled them to capture the complexity in some stages of development in second language writings from a Dynamic Systems perspective. Finite Verb Ratio (FVR) and Average Word Length (AWL) were mentioned as the best broad syntactic and lexical measures to employ in the research which does not focus on specific constructions, but more on the averages of many instances. The difference is that FVR also deals with the internal complexification of clauses.

Other four specific complexity measures were also employed by Verspoor et al. (2017) in their research on linguistic complexity in academic writing to obtain further tracing of development at different proficiency levels; finite adverbial, nominal and relative clauses, and non-finite constructions. There were some noticeable shifts reflected in a competitive or supportive relationship, and in positive or negative correlation between variables being measured. It then came to the conclusion that that learners have their own individual learning paths (Verspoor et al., 2017). To emphasize, Chan et al. (2015) postulated that even for identical twins who got the same
amount of exposure to a language and had the same teacher, when they were even examined generally on both speaking and writing skills, they still performed different developmental patterns.

Next, in another related research, Verspoor et al. (2012) gathered 437 texts written by Dutch learners of English as a foreign language between the ages of 11 and 14 with similar scholastic aptitude scores to seek whether proficiency levels might affect their dynamic second language developmental patterns on writings. A multitude of variables was included to generate significant data in addressing variability, among others sentence length, the Guiraud Index, all dependent clauses combined, all chunks combined, all errors combined, and the use of present and past tense. In the end, it revealed that there was a non-linear development, and the relationships among the variables almost always did change.

In addition, Spoelman and Verspoor (2010) also showed their efforts to prove that specific change occurred in the interaction among different complexity measures. The growers (word complexity, sentence complexity, noun phrase (NP) complexity) were seen connected to each other, and thus, they developed at exactly the same time. The accuracy rate was also measured to see whether variability might appear if complexity measures were involved. To conclude, this study strengthens the idea that the interaction within these subsystems of accuracy and complexity was found to provide useful insights into the second language developmental process.

Those previous studies underlie this study to delve into the lexical complexity of L2 learner’s academic writings. Therefore, this study attempts to analyze Lexical Diversity (LD) and Lexical Sophistication (LS) and their interaction as well. Vocabulary Diversity (VocD) and Average Word Length in Morphemes (AWLIM) are two key measures in helping the author to visualize the patterns in L2 development. The hypotheses of this study are as follows:

1. LD and LS will present a dynamic trajectory in the learner’s writing development over time, not to mention his proficiency level also increased within the writing process.
2. The dynamic interactions between LD and LS will be found during the tracing of variability.

3. METHODS

This study analyzes the intra-individual variability using longitudinal data, particularly written academic texts, of a learner over five years (2010-2015). The Dynamic Systems Theory (DST) is perceived to be able to contribute to visualizing the developmental process. The emerging variability will be investigated to show whether there is any dynamic interaction between Lexical Diversity (LD) and Lexical Sophistication (LS).

3.1 Subject

The subject of this study was a 25-year-old Indonesian Master’s student majoring in Applied Linguistics at a university in the Netherlands. He was first officially exposed to English as a foreign language since he was in his fourth year (around 9 years old) in elementary school, so up until 2015, he had been learning
English for 16 years approximately. He already took four English proficiency tests (e.g., TOEFL ITP in 2011, 2012, 2013, and IELTS in 2015), which showed that he was at a near-advanced level as a learner of English. He had studied and lived in the Netherlands for one year, in which English is omnipresent, and thus, he was much exposed to a large amount of English from the environment, lectures, classmates, people, media, and other sources.

3.2 Material

Twelve pieces of academic writings were chosen as the corpus to analyze. They were written from the age of 20 to 25 under various kinds of topics, for instances, literature, journalism, culture, contemporary issues, education, linguistics, language, and plagiarism. Due to the different lengths of each writing sample, instead of taking them wholly, averagely only 200 words per text were selected from the first one or two paragraphs so that it could be fairly judged and examined. Additionally, choosing only 200 words per text was considered sufficiently representative to provide initial insight into the dynamic development of the participant’s writings.

Average Word Length in Morphemes (AWLiM) and Vocabulary Diversity (VocD) were two lexical complexity measures used to detect the presence of variability in the developmental process generally, in which both measures were relevant for the purpose of this study. VocD represents the various words by taking random subsamples from the text and calculating a TTR (Type Token Ratio). The higher its value of VocD, the more varied the text. Another variable, AWLiM, indicates the word length in morphemes on average, which was considered useful in this study and related to VocD as well, in which if the learner deployed more (various) morphemes, the vocabulary might be more diverse, so that the interaction between two growers could be possibly revealed.

3.3 Design and Analyses

There were several methods and tools employed to explain the variability in developmental time-series data. At first, 12 texts (200 words per text) were coded and analyzed using CLAN (Computerized Language Analysis) program, a supporting tool designed and written by Leonid Spektor (MacWhinney, 2000). The variables being measured were Vocabulary Diversity (VocD) and Average Word Length in Morphemes (AWLiM). Moreover, to obtain the general picture of the developmental patterns, the variables were then plotted, and the trend lines were added using a second-degree polynomial. Next, one of the tools developed by van Geert and van Dijk (2002), namely moving min-max graphs, was used to inspect the variability. Since there were only 12 short-writing samples, the author preferred to select three of the minimum and maximum values to be processed by this tool. In addition, the raw data was also detrended, so that the inclining slope would not disturb the variability and interactional degree (Verspoor et al., 2011). After that, the resampling technique and Monte Carlo Analyses were used to identify whether the peaks were just coincidental or patterned and whether there was any significant difference between the variables. The tool created by Steinkrauss (2016) was also used. The data was then reshuffled for 5000 times with the significance level of 0.05, so the p-value would be considered significant if the simulation revealed that the peaks did not reach more than 250 times.
The next four steps in looking at the interactions among variables from a Dynamic Systems perspective followed van Dijk et al. (2011): 1) visual inspection, 2) smoothing (LOESS/LOWESS – Logically Weighted Scatterplot Smoothing); a tool developed by Peltier (2009), 3) normalizing (0-1 scaling), and 4) simple and moving correlation. Finally, the findings were interpreted to give a clear conclusion of the variability in the dynamic development in academic writing of a learner of English.

4. RESULTS AND DISCUSSION

This section presents the results after observing carefully the developmental data in terms of two specific variables, namely Lexical Diversity/LD (Vocabulary Diversity/VocD) and Lexical Sophistication/LS (Average Word Length in Morphemes/AWLiM). The changing interactions between the growers that happened in the learner’s writings were also drawn as a comparison.

4.1 Lexical Diversity (LD) – Vocabulary Diversity (VocD)

In this study, LD was measured by providing the value of VocD. Figure 1 depicts the VocD graph with a 2nd degree polynomial trend line to discover information about the variability patterns.

![Figure 1: The development of vocabulary diversity (VocD) including a 2nd degree polynomial trend line in 12 academic writings over five years.](image)

Looking at the line-graph in Figure 1, there were some moments of progress and regress in the development of VocD, in which Text 12 had the highest peak, whereas the lowest trough was showed in Text 6. The growth trend line using a 2nd degree polynomial moved upward in the end after having several downward situations since the learner also increased his language proficiency level over time. There was also an extreme incline from Texts 6 to 7 continued with an extreme decline from Texts 7 to 8. The detrended raw data of VocD was then created so that the variability could be signified in Figure 2.
Figure 2. The detrended raw data (residual values) of the VocD development in 12 academic writings over five years.

The detrended raw data was able to visualize the presence of variability better than the un-detrended one. It can be seen that the variability already oscillated from the first until the last text samples. Moreover, from the graph, Text 7 showed the highest top, while the trough went very low in Text 6. Figure 3 plotted a min-max graph which was helpful to see whether the clear patterns in which peaks and dips occurred can be identified.

Figure 3. The min-max graph of the VocD development in 12 academic writings over five years.

From Figure 3, it can be seen that since the beginning, the bandwidth of scores did not remain stable at all. It always moved upward or downward unpredictably. The range of scores increased extremely from Text 6, but it soon went down again, until it
showed quite steady upward growth movements from Text 8 onward. In other words, the variability did not stay at the same level of bandwidth after all.

Even though there were some large peaks in the development of VocD, the $p$-value obtained after performing resampling and Monte Carlo analyses affirmed that the peaks were not significant ($p=0.1992$). Among 5000 times in the simulation, they occurred as many as 996 times.

### 4.2 Lexical Sophistication (LS) – Average Word Length in Morphemes (AWLiM)

Average Word Length in Morphemes (AWLiM) was measured to indicate the degree of Lexical Sophistication (LS) in the learner’s academic writings over five years. Figure 4 plots a 2\textsuperscript{nd} degree polynomial trend line of the AWLiM development.

![Figure 4](image.png)

**Figure 4.** The development of average word length in morphemes (AWLiM) including a 2\textsuperscript{nd} degree polynomial trend line in 12 academic writings over five years.

Figure 4 shows that, comparing to the previous variable, the developmental patterns of AWLiM were more stable, suggesting that the learner used longer and more morphemes on average over five years. The highest peak was seen in Text 11, and Text 4 showed the lowest dip among all data points. To display the variability trajectory in a clearer graph, the data was detrended and showed in Figure 5.
Figure 5. The detrended raw data (residual values) of the AWLiM development in 12 academic writings over five years.

The graph in Figure 5 indicates that initially, the data points stayed in a quite steady position, but the variability then happened here and there, such as from Texts 2 to 4, 5 to 7, 8 to 10, and 10 to 12. Text 11 held the highest summit, almost the same as Text 3. Figure 6 provides the min-max graph to give further information on the range of variability being observed.

Figure 6. The min-max graph of the AWLiM development in 12 academic writings over five years.

The upward movement can be seen from the graph in Figure 6. After some moments of progress and regress, the degree of variability was found to be at the same level starting from Text 9 onward. However, since the $p$-value was 0.1876 and the occurrence was of 938 times, the peaks were considered not significant and perceived just as a coincidence.
4.3 Interaction between Vocabulary Diversity (VocD) and Average Word Length in Morpheme (AWLiM)

To inspect the presence of dynamic interaction between Vocabulary Diversity (VocD) and Average Word Length in Morpheme (AWLiM), Figure 7 plots together with the two growers after being rescaled to the range of 0 to 1. This was done to make sure that those data points can be compared fairly.

The trend lines shown in the graphs depict upward growth patterns between the two growers. Initially, each grower had their own trajectory which was against each other. For instance, in Text 1, when VocD moved up, AWLiM moved down, instead. Nevertheless, they displayed relatively parallel movements afterward even until the last text, which led to supportive interaction. It was also emphasized by a Pearson correlation, stating that the correlation between VocD and AWLiM was positive, strong, and significant ($r=0.76$, $p<0.05$; two-tailed). In addition, to capture the general tendencies between the variables and how they relate to each other, the data were smoothed and normalized and both were plotted in Figure 8.
At first, the relation between the two growers seemed to be negative, but it then changed to be in a positive correspondence, ending up in a similar movement simultaneously. A moving window of correlation was then used and plotted in the following figure to provide more precise information on the dynamic interaction of VocD and AWLiM.

The change of interaction from competitive (negative) to supportive (positive) emerged from the second or third text onward. Although there were some dips in Texts 5 and 8, they were considered incidental. To highlight the periods of this interaction, Figure 10 was created.
Figure 10. Moving correlation (three data points) including highlights of interactional periods between the growers.

The periods where the interactions took place were highlighted in the graph above. The highlighted ones were some strong and positive correlations, while the one below was showed when the two growers initially had their first interaction. By doing this, the dynamic developmental patterns can be well-observed as the emerging variability is noticeable to scrutinize.

4.4 Discussion

This study aims at addressing whether intra-individual variability of Lexical Diversity (LD), in this case, Vocabulary Diversity (VocD), and Lexical Sophistication (LS), specifically Average Word Length in Morphemes (AWLiM), can provide important insights on (second) language development. The study also attempts to investigate the presence of dynamic interaction between the two growers. The findings of this study indicate that the learner’s academic writings go through a dynamic developmental process as what the Dynamic Systems Theory (DST) posits its arguments.

There are three main points to reveal that this study is in relation to DST. Firstly, although the interaction between the variables does change in various ways, several attractor states can still be recognized clearly. This emphasizes the notion in DST that during the developmental process, variability does not stay stagnantly. There is a time where there is little change over the period of time like what happened in Texts 5 and 6. DST claims that this variability is the source of development (van Dijk, 2003 in de Bot et al., 2005). Secondly, the (second) language development in the learner’s academic writing is considered non-linear and unpredictable. It is due to internal and external factors to take into account in the developmental patterns. When a language learner becomes more advanced, it does not necessarily mean that his writing skill will always show the upward growth movement. There must be a moment, although the proficiency level increases, when his writing does not develop well. Texts 5 and 6, for instances, indicate the decline in the process of development. The internal factor, such as motivation, might have an impact on that. Thirdly, it was found that the variables
being measured did change their interaction from competitive to supportive, slightly back to be competitive, but eventually, the interaction tends to be supportive. This phenomenon is in line with DST, which says that self-organization in a language system does exist. From the moving correlation, it can be inferred that the interaction of variables is complex and predominantly iterates, in which it reorganizes itself to have parallel movements on one another.

5. CONCLUSION

The present study has revealed interesting insights into the developmental pattern of learner’s academic writings. By investigating the intra-individual variability of Lexical Diversity (Vocabulary Diversity or VocD) and Lexical Sophistication (Average Word Length in Morphemes or AWLiM), it can be concluded that the learner’s academic writings experience a dynamic developmental process, which is in line with the Dynamic Systems Theory.

This study cannot be just generalized to other studies as it focuses only on the intra-individual variability of VocD and AWLiM. Another limitation of this study is the use of longitudinal data from only one learner. Further research on variability integrated with learner’s individual differences, such as age, gender, proficiency level, and motivation, from a larger population or sample size might be interesting to conduct. In addition, having more academic writing samples written over a longer period is necessary to investigate learners’ dynamic second language developmental patterns thoroughly. This is considered to be important to help learners be aware of their unique language learning trajectory, especially of writing skills, so that they can maintain a more stable linguistic development in the future.

REFERENCES


http://peltiertech.com/loess-utility-awesome-update/

https://doi.org/10.1093/applin/amq001

http://www.let.rug.nl/sldmethods/


https://doi.org/10.1016/S0163-6383(02)00140-6


https://doi.org/10.4000/rdlc.1450

https://doi.org/10.1111/j.1540-4781.2008.00715.x

https://doi.org/10.1016/j.jslw.2012.03.007